

Lehrerpersönlichkeit und Unterrichtsqualität aus Schülerperspektive

Jessika Bertram





Erziehungswissenschaft

Lehrerpersönlichkeit und Unterrichtsqualität aus Schülerperspektive

– unter Berücksichtigung des individuellen Fachinteresses

Inaugural-Dissertation

zur Erlangung des Doktorgrades Dr. phil.

im Fachbereich Erziehungswissenschaft und Sozialwissen-
schaften

der Westfälischen Wilhelms-Universität in Münster

vorgelegt von

Jessika Sarah Bertram, geb. Haßio

-2017-

Dekan des Fachbereichs Erziehungswissenschaft und Sozialwissenschaften: Prof. Dr. Martin Bosen

Erste Gutachterin: Prof. Dr. Sabine Gruehn
(Westfälische Wilhelms-Universität Münster)

Zweite Gutachterin: Prof. Dr. Sylvia Rahn
(Bergische Universität Wuppertal)

Tag der Promotion: 27.02.2018

Jessika Bertram

**Lehrerpersönlichkeit und Unterrichtsqualität aus
Schülerperspektive**



Wissenschaftliche Schriften der WWU Münster

Reihe VI

Band 18

Jessika Bertram

Lehrerpersönlichkeit und Unterrichtsqualität aus Schülerperspektive

Wissenschaftliche Schriften der WWU Münster

herausgegeben von der Universitäts- und Landesbibliothek Münster

<http://www.ulb.uni-muenster.de>



Bibliografische Information der Deutschen Nationalbibliothek:

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

Dieses Buch steht gleichzeitig in einer elektronischen Version über den Publikations- und Archivierungsserver der WWU Münster zur Verfügung.

<http://www.ulb.uni-muenster.de/wissenschaftliche-schriften>

Jessika Bertram

„Lehrerpersönlichkeit und Unterrichtsqualität aus Schülerperspektive“

Wissenschaftliche Schriften der WWU Münster, Reihe VI, Band 18

Verlag readbox publishing GmbH – readbox unipress, Münster

<http://unipress.readbox.net>

Zugl.: Diss. Universität Münster, 2018

Dieses Werk ist unter der Creative-Commons-Lizenz vom Typ 'CC BY 4.0 International'

lizenziert: <https://creativecommons.org/licenses/by/4.0/deed.de>

Von dieser Lizenz ausgenommen sind Abbildungen, welche sich nicht im Besitz der Autorin oder der ULB Münster befinden.



ISBN 978-3-8405-0207-1

(Druckausgabe)

URN urn:nbn:de:hbz:6-25149472578

(elektronische Version)

direkt zur Online-Version:

© 2019 Jessika Bertram

Satz:

Jessika Bertram

Titelbild:

Vasylysk (Graffiti in Nanyuki). CC BY-SA 3.0

https://commons.wikimedia.org/wiki/File:Big_Five_Game.jpg

Umschlag:

ULB Münster



Inhaltsverzeichnis

Abbildungsverzeichnis	v
Tabellenverzeichnis.....	vii
1 Einleitung.....	1
1.1 Perspektivität und Wahrnehmung.....	5
1.2 Aufbau der Arbeit	12
2 Die gute Lehrkraft als Synthese von Persönlichkeit und Unterrichts- qualität.....	17
2.1 Operationalisierung des Konstrukts der guten Lehrkraft.....	23
2.2 Forschungsüberblick zur guten Lehrkraft.....	27
2.2.1 Forschungsbefunde zur Lehrerpersönlichkeit	35
2.2.2 Dr.Fox-Effekt.....	41
2.2.3 Lehrer aus Schülersicht.....	42
2.2.4 Lehrerpersönlichkeit – Selbsteinschätzungen vs. Fremd- einschätzungen.....	46
2.2.5 Assessment-Center.....	47
3 Dialektik der Persönlichkeit – Schnittstelle zwischen Innen- und Außenwelt	53
3.1 Forschungslinien und Forschungstraditionen der Persönlichkeitsforschung	60
3.2 Operationalisierung des Konstrukts Persönlichkeit.....	61
3.3 BigFive der Persönlichkeit.....	65
3.3.1 Entwicklung der BigFive	66
3.3.2 Systematik des BigFive-Konstrukts: Dimensionen/Facetten	69
3.4 Forschungsüberblick zur Persönlichkeit.....	72
3.4.1 Das Verhältnis von Fremd- und Selbsteinschätzung der Persönlichkeit.....	73
3.4.2 Kontextspezifität.....	79
3.4.3 Weitere Forschungsbefunde	80
4 Diskurs über normative Ausrichtungen der Unterrichtsqualität.....	83
4.1 Forschungslinien und Forschungstraditionen zur Unterrichtsqualität....	87
4.2 Operationalisierung des Konstrukts Unterrichtsqualität.....	91

Inhaltsverzeichnis

4.3	Prozessmerkmale der Unterrichtsqualität.....	94
4.3.1	Entwicklung von Prozessmerkmalen des Unterrichts.....	94
4.3.2	Systematik der Prozessmerkmale des Unterrichts	96
4.4	Forschungsüberblick zur Unterrichtsqualität	98
4.4.1	Schülerperzipierte Unterrichtsqualität	100
4.4.2	Fachspezifität.....	106
5	Ableitung von Forschungsdesiderata und Fragestellungen.....	109
6	Datengrundlage und Untersuchungsdesign	117
6.1	Anlage der DFG-Studie Determinanten des Schülerfeedbacks	117
6.2	Datenerhebung.....	118
6.3	Stichprobenbeschreibung der verwendeten Daten	120
6.3.1	Schülerstichprobe	120
6.3.2	Lehrerstichprobe.....	121
6.3.3	Gemeinsame Stichprobe von Lehrern und Schülern.....	122
7	Umgang mit der Mehrebenenstruktur der Daten.....	123
8	Umgang mit fehlenden Werten.....	127
9	Interpretation der Modellgüte.....	131
10	Erhebungsinventar und Validierungsverfahren	137
11	Lehrerpersönlichkeit: Instrument	143
11.1	BFI-S zur Erhebung der Lehrerpersönlichkeit	143
11.2	Beschreibung der einzelnen Skalen resp. Dimensionen.....	145
11.2.1	Gewissenhaftigkeit	148
11.2.2	Extraversion.....	149
11.2.3	Offenheit für Erfahrungen	150
11.2.4	Resilienz resp. Neurotizismus	151
11.2.5	Verträglichkeit.....	152
11.3	Validierung des Persönlichkeitskonstrukts: Messmodelle der Schülerantworten	153
11.3.1	Explorative Faktorenanalyse (EFA).....	154
11.3.2	Vier-Modell-Prüfung.....	157
11.3.3	Reliabilität der einzelnen Skalen resp. Dimensionen.....	159

Inhaltsverzeichnis

11.3.4	Mehrebenenmodelle der fachspezifischen Gesamt- konstrukte	161
11.4	Messmodelle der Lehrerantworten	163
11.4.1	Reliabilität der einzelnen Skalen resp. Dimensionen	163
11.4.2	Gesamtkonstrukt	165
12	Unterrichtsqualität: Instrument	167
12.1	Prozessmerkmale der Unterrichtsqualität	167
12.2	Beschreibung der einzelnen Skalen resp. Prozessmerkmale des Unterrichts.....	169
12.2.1	Angemessenes Interaktionstempo	170
12.2.2	Angemessene Methodenvariation.....	171
12.2.3	Disziplin.....	171
12.2.4	Störungsprävention	172
12.2.5	Erklärkompetenz.....	173
12.2.6	Motivierung	174
12.2.7	Alltagsbezug	175
12.3	Validierung des Konstrukts der Unterrichtsqualität	176
12.3.1	Vier-Modell-Prüfung	177
12.3.2	Reliabilität der einzelnen Skalen	180
12.3.3	Mehrebenenmodelle der fachspezifischen Gesamt- konstrukte.....	181
13	Analysestrategien	183
13.1	Forschungsfrage 1: Welcher Zusammenhang besteht zwischen selbst- und fremdwahrgenommener Lehrerpersönlichkeit?	183
13.2	Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten Lehrerpersönlichkeit und Unterrichtsqualität?.....	184
14	Deskriptive Befunde.....	191
14.1	Selbsteinschätzung der Lehrerpersönlichkeit	192
14.1.1	Fachübergreifende Deskription	193
14.1.2	Fachspezifische Deskription.....	195
14.1.3	Interkorrelationen zwischen den von den Lehrkräften selbsteingeschätzten BigFive-Dimensionen	198
14.2	Schülereinschätzung der Lehrerpersönlichkeit.....	199

Inhaltsverzeichnis

14.2.1	Deutsch.....	201
14.2.2	Mathematik.....	202
14.2.3	Geschlechterdifferenzen innerhalb der Fächer.....	203
14.2.4	Interkorrelationen zwischen den schülerperzipierten BigFive-Dimensionen der Lehrkräfte	203
14.3	Unterrichtsqualität aus Schülersicht.....	205
14.3.1	Deutsch.....	206
14.3.2	Mathematik.....	208
14.3.3	Interkorrelationen zwischen den schülerperzipierten Prozessmerkmalen der Unterrichtsqualität	209
15	Forschungsfrage 1: Welcher Zusammenhang besteht zwischen selbst- und fremdwahrgenommener Lehrerpersönlichkeit?	213
15.1	Zusammenhang von verschiedenen Wahrnehmungsperspektiven der Lehrerpersönlichkeit – Korrelationen	214
15.2	Kontrolle des Lehrkraftgeschlechts im Fach Mathematik	217
16	Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten Lehrerpersönlichkeit und Unterrichtsqualität?	219
16.1	Perspektivenübergreifender Zusammenhang von Persönlichkeits- selbsteinschätzungen der Lehrkräfte mit schülerperzipierten Pro- zessmerkmalen des Unterrichts – Korrelationen und Interpretation...	220
16.2	Perspektivenspezifischer Zusammenhang von Lehrer- persönlichkeit und Unterrichtsqualität – Regressionsanalysen.....	223
16.2.1	...im Fach Deutsch.....	229
16.2.2	...im Fach Mathematik.....	232
17	Interpretation der ermittelten Forschungsbefunde.....	237
17.1	Interpretation des Vorbefunds	238
17.2	Interpretation der korrelativen Befunde	239
17.3	Interpretation der regressionsanalytischen Befunde	243
18	Zusammenfassende Diskussion.....	249
19	Fazit	255
20	Limitationen und Ausblick	261
	Literaturverzeichnis	263
	Anhang.....	303

Abbildungsverzeichnis

<i>Abb. 1: Zusammenhänge zwischen Persönlichkeit und Verhalten von Lehrkräften (aus Mayr/Neuweg 2006, S. 191).</i>	38
<i>Abb. 2: Modell nach Costa/McCrae 1989.</i>	65
<i>Abb. 3: Angebots-Nutzungs-Modell des Unterrichts (Helmke 2014, S. 71)</i>	91
<i>Abb. 4: Generalfaktormodell</i>	138
<i>Abb. 5: genestetes Modell</i>	138
<i>Abb. 6: fächergetrenntes Modell unter Zulassung von Korrelationen</i>	139
<i>Abb. 7: fächergetrenntes Modell</i>	139
<i>Abb. 8: Gesamtmodell auf Individualebene.</i>	141
<i>Abb. 9: Gesamtmodell auf Individual- sowie Klassenebene.</i>	141
<i>Abb. 10: Gesamtmodell auf Individualebene und einem Generalfaktor auf Klassenebene.</i>	141
<i>Abb. 11: BFI-S gemäß SOEP (vgl. Gerlitz/Schupp 2005)</i>	145
<i>Abb. 12: vereinfachtes Modell über Korrelationen zwischen den verschiedenen Lehrer-BigFive-Dimensionen aus Lehrer- und Schülersicht.</i>	184
<i>Abb. 13: saturiertes Strukturmodell mit Effekten der Lehrerpersönlichkeit auf die Unterrichtsqualität (Modell 1a).</i>	185
<i>Abb. 14: Strukturmodell mit Effekten der Lehrerpersönlichkeit auf die Unterrichtsqualität mit Korrelationen zwischen sehr hoch korrelierenden Prädiktoren (Modell 2a).</i>	187
<i>Abb. 15: Strukturgleichungsmodell mit Effekten der Lehrerpersönlichkeit (als Faktor 2. Ordnung) auf die Unterrichtsqualität (Modell 3).</i>	187
<i>Abb. 16: Fachübergreifende Mittelwerte der Lehrer-BigFive aus Lehrersicht (Selbsteinschätzung)</i>	193
<i>Abb. 17: Fachübergreifende Verteilung der Lehrkräfte-BigFive (Selbsteinschätzung)</i>	194
<i>Abb. 18: Mittelwerte der BigFive-Selbsteinschätzung von Deutsch- und Mathematiklehrkräften im Vergleich</i>	196
<i>Abb. 19: Verteilungen der Lehrkräfte-BigFive (Selbsteinschätzung) im Fach Deutsch</i>	197

Abbildungsverzeichnis

<i>Abb. 20: Verteilungen der Lehrkräfte-BigFive (Selbsteinschätzung) im Fach Mathematik.....</i>	<i>198</i>
<i>Abb. 21: Mittelwerte der BigFive der Deutsch- und Mathematiklehrkräfte aus Schülersicht (Fremdeinschätzung)</i>	<i>200</i>
<i>Abb. 22: Häufigkeitsverteilung der Deutschlehrkräfte-BigFive aus Schülersicht (Fremdeinschätzung)..</i>	<i>201</i>
<i>Abb. 23: Häufigkeitsverteilung der Lehrkräfte-BigFive aus Schülersicht (Fremdeinschätzung).</i>	<i>202</i>
<i>Abb. 24: Mittelwerte der Prozessmerkmale der Unterrichtsqualität aus Schülersicht im Fach Deutsch und im Fach Mathematik</i>	<i>206</i>
<i>Abb. 25: Häufigkeitsverteilung der Prozessmerkmale des Deutschunterrichts aus Schülersicht</i>	<i>207</i>
<i>Abb. 26: Häufigkeitsverteilung der Prozessmerkmale des Mathematikunterrichts aus Schülersicht</i>	<i>209</i>
<i>Abb. 27: Überblick über Modell 1b im Fach Deutsch – Effekte von Schülereinschätzungen der Lehrerpersönlichkeit auf die Wahrnehmung der Unterrichtsqualität auf Individualebene und auf Klassen resp. Lehrkraftebene</i>	<i>229</i>
<i>Abb. 28: Überblick über Modell 1b im Fach Mathematik – Effekte von Schülereinschätzungen der Lehrerpersönlichkeit auf die Wahrnehmung der Unterrichtsqualität auf Individualebene und auf Klassen resp. Lehrkraftebene</i>	<i>232</i>

Tabellenverzeichnis

<i>Tab. 1: Korrelationen zwischen Selbsteinschätzung und aggregierten Schülereinschätzungen (nach Mayr/Neuweg 2006, Signifikanzen ergänzt durch Mayr 2019)</i>	<i>46</i>
<i>Tab. 2: BigFive-Dimensionen mit ihren Facetten (adaptiert aus Ostendorf/Angleitner 2003, ergänzt anhand von Simon 2006).</i>	<i>70</i>
<i>Tab. 3: Angestrebte Sollstichprobe der durchgeführten Studie</i>	<i>118</i>
<i>Tab. 4: Netto-Schülerstichprobe der durchgeführten Studie</i>	<i>120</i>
<i>Tab. 5: Netto-Lehrerstichprobe der durchgeführten Studie.....</i>	<i>121</i>
<i>Tab. 6: Stichprobe der Synthese aus Netto-Schüler- und -Lehrerdaten</i>	<i>122</i>
<i>Tab. 7: Fit-Indizes nach Schermelleh-Engel/Moosbrugger/Müller 2003, S. 38; Werte in Klammern nach Hu & Bentler 1999 (siehe auch Zinnbauer/Eberl 2004).....</i>	<i>134</i>
<i>Tab. 8: In dieser Untersuchung eingesetztes Instrument (BFI-S)</i>	<i>147</i>
<i>Tab. 9: Explorative Faktorenanalyse der Lehrer-BigFive aus Schülersicht.....</i>	<i>155</i>
<i>Tab. 10: Messmodelle der Persönlichkeitsdimensionen</i>	<i>158</i>
<i>Tab. 11: Deutsch-Mathematik-Korrelationen (manifest auf Individualebene) innerhalb der einander entsprechenden Dimensionen.....</i>	<i>159</i>
<i>Tab. 12: Skalenwerte der erfassten BigFive-Dimensionen</i>	<i>160</i>
<i>Tab. 13: (Mehrebenen-)Messmodelle im Vergleich</i>	<i>162</i>
<i>Tab. 14: Cronbachs α der eingesetzten Skalen zur Erhebung der Lehrer-BigFive im Selbstbericht</i>	<i>164</i>
<i>Tab. 15: Messmodelle der Lehrerantworten zu den erhobenen Persönlichkeitsdimensionen.....</i>	<i>166</i>
<i>Tab. 16: Eingesetzte Skalen zur Erfassung der Prozessqualität des Unterrichts</i>	<i>169</i>
<i>Tab. 17: Messmodelle der schülerperzipierten Prozessmerkmale des Unterrichts</i>	<i>178</i>

Tabellenverzeichnis

<i>Tab. 18: Deutsch-Mathematik-Korrelationen (manifest auf Individualebene) innerhalb der einander entsprechende Dimensionen</i>	180
<i>Tab. 19: Reliabilität der Prozessmerkmale aus Schülersicht</i>	180
<i>Tab. 20: modifizierte (Mehrebenen-)Messmodelle der Prozessmerkmale</i>	181
<i>Tab. 21: Interkorrelationen der BigFive-Dimensionen aus Lehrersicht (Selbsteinschätzung)</i>	199
<i>Tab. 22: Interkorrelationen der BigFive-Dimensionen aus Schülersicht</i>	204
<i>Tab. 23: Interkorrelationen zwischen den Prozessmerkmalen der Unterrichtsqualität nach Fächern.....</i>	210
<i>Tab. 24: Standardisierte Korrelationen zwischen den Lehrerselbsteinschätzungen und den Schülereinschätzungen (Fremdeinschätzung) der BigFive-Persönlichkeitsdimensionen</i>	215
<i>Tab. 25: Interkorrelationsmatrix der schülerperzipierten und der selbsteingeschätzten Lehrkraft-BigFive-Dimensionen auf Lehrkraftebene in Deutsch/Mathematik</i>	216
<i>Tab. 26: Standardisierte Korrelationskoeffizienten von Schülerfremdeinschätzung der Lehrerpersönlichkeit mit Lehrkräfteselbsteinschätzung im Fach Mathematik unter Einbezug des Lehrkraftgeschlechts als Moderatorvariable.....</i>	217
<i>Tab. 27: Korrelationen der BigFive-Selbsteinschätzungen der Lehrerinnen und Lehrer in Deutsch/Mathematik mit den schülerperzipierten Prozessmerkmalen des Unterrichts; manifest und auf Lehrkraftebene aggregiert</i>	221
<i>Tab. 28: Güteindizes der Strukturmodelle in Deutsch und Mathematik</i>	225
<i>Tab. 29: Modell 1b – signifikante Effekte (Regressionskoeffizienten) der Lehrerpersönlichkeit auf die Unterrichtsqualität.</i>	226
<i>Tab. 30: Modell 3 – standardisierte Regressionskoeffizienten und Varianzaufklärung der Prozessmerkmale des Unterrichts durch einen übergeordneten Persönlichkeitsfaktor 2. Ordnung</i>	227

1 Einleitung

Mit Foucault könnten bisher gängige Ansätze der Schul- und Unterrichtsforschung zur Bedeutung der Lehrperson für die Gestaltung des Unterrichts und der erreichten Unterrichtsqualität als Reproduktion bestehender Machtverhältnisse angesehen werden, da die Deutungshoheit über die Person in der Regel als in dieser selbst angelegt aufgefasst wird (vgl. Sternfeld 2010). Zwar sind Machtverhältnisse keineswegs Thema dieser Arbeit, doch führt offenbar genau diese immer noch bestehende Zuschreibung der Deutungshoheit zu dem stets gleichen Ansatz innerhalb der Forschung: die im Selbstbericht erhobene Lehrerpersönlichkeit mit der – hier ist die Schul- und Unterrichtsforschung hinsichtlich des Einbezugs unterschiedlicher Perspektiven bereits weitaus fortschrittlicher – mittlerweile meist aus Schülerperspektive erhobenen Unterrichtsqualität¹ zu verknüpfen. Allerdings bleibt dabei in weiten Teilen unberücksichtigt, dass die Lehrerin oder der Lehrer immer in einem Verhältnis zu ihren/seinen Schülerinnen und Schülern steht, dem ein Wechselspiel der Wahrnehmungen zugrunde liegt. So, wie die Lehrkraft ihre Schülerinnen und Schüler in einem anderen Licht sieht als diese sich selbst, sehen logisch stringent auch die Schülerinnen und Schüler ihre Lehrerinnen und Lehrer anders, als sich diese selbst begreifen. Daraus leitet sich ab, was bereits Clausen (2002) mit seiner Forschung in Bezug auf die Unterrichtsqualität herausgestellt hat: dass Wahrnehmung innerhalb des Unterrichts perspektivenspezifisch stattfindet, ohne der einen oder der anderen Perspektive dadurch Gültigkeit ab-, jedoch auch ohne eine Deutungshoheit zuzuschreiben. Diese Perspektivenspezifität betrifft allerdings ebenso wie ihren Unterricht auch die Persönlichkeit von Lehrerinnen und Lehrern, wodurch es angebracht scheint, diese beiden Konstrukte auch entsprechend aus Schülersicht zu erheben und innerhalb dieser Perspektive in Zusammenhang zu bringen, sofern man eine Bedeutsamkeit der Lehrerpersönlichkeit für die Unterrichtsqualität vermutet. Bislang hat das in der Schul- und Unterrichtsforschung jedoch nicht überzeugend stattgefunden. Ein möglicher Grund dafür mag in der schier unübersichtlichen

¹ Dass die Schülerinnen und Schüler in der Lage sind, den Unterricht, an dem sie teilnehmen, auch valide einzuschätzen, ist eine noch eher junge Erkenntnis, die sich jedoch schnell durchsetzen konnte; v.a. nachdem gezeigt werden konnte, dass solche Einschätzungen durchaus relevante Hinweise für deren Lernleistungen geben können (vgl. Gruehn 2000).

Vielfalt an Ansätzen, Informationen und Forschungen zu den Konstrukten liegen, die die *gute* Lehrperson als Synthese von Persönlichkeit und Unterrichtsqualität abbilden: Die Eingaben „Persönlichkeit“ und „personality“ ergeben zusammengenommen fast 2000 Suchergebnisse beim psychologischen Suchportal *Psychspider* (Stand 05.04.2017), die eher umgangssprachliche, persönlichkeitsbezogene Eingabe „Ich bin ein Mensch, der...“ bei *Google* sogar rund 50 Millionen (Stand 05.04.2017). Hinzu kommen die 13.000 Suchergebnisse, die unter dem Begriff „Unterrichtsqualität“ in der Suchdatenbank *Fachportal Pädagogik* (Stand 05.04.2017) erscheinen. Dieser Befund fordert eine strikte Filterung der für den Zweck notwendigen Informationen, was ebenso innerhalb der vorliegenden Arbeit erfolgt.

Ein kurzer Überblick über die Konstrukte soll nun in das Anliegen und die Fragen, die dieser Arbeit zugrunde liegen, einleiten. Danach folgt eine Erläuterung der grundsätzlichen Ausrichtung dieser Arbeit (in Kap. 1.1) und erst anschließend wird der weitere Aufbau der Arbeit vorgestellt (Kap. 1.2).

Unterrichtsqualität lässt sich kurz und knapp zunächst als multidimensionales Konstrukt begreifen, das sich jeweils entlang verschiedener Paradigmen definiert. Diese Paradigmen rahmen das Verständnis von Unterrichtsqualität sowohl inhaltlich als auch normativ – sogenannter *guter* Unterricht ist dann immer das, was entsprechend des jeweiligen Paradigmas als Resultat sichtbar wird. In der Regel bestehen Sinn und Zweck *guten* Unterrichts darin, eine entsprechend der gesetzten Norm nachweisbare Wirkung bei den Schülerinnen und Schülern zu erzielen. Demnach ist das theoretische Konstrukt der Unterrichtsqualität eng an seine empirische Erfassung geknüpft: beide entwickeln sich in wechselseitiger Weise parallel.

Die Lehrerpersönlichkeit stellt darüber hinaus die Synthese des Lehrers oder der Lehrerin als Person und dessen resp. deren Unterrichtsqualität dar. Die Bedeutung der Lehrpersonen – derer sich dieselben laut einer Studie des Allensbach-Instituts (2012) wenig bewusst sind – kommt in etablierten Modellen zur Unterrichtsqualität stets zum Tragen, in der Regel jedoch ohne näher definiert zu werden. So veranschaulichen bspw. Helmkes Makromodell der Bedingungsfaktoren schulischer Leistungen (2014, S. 28) sowie seine diversen Angebot-Nutzungs-Modelle (2004, 2007, 2009, 2014) zwar die Bedeutsamkeit der Lehrerpersönlichkeit bzw. (scheinbar wahllos) ausgesuchter Persönlichkeitseigenschaften bzw. -merkmale. Jegliche differenzierte, theoretisch fundierte Auseinandersetzung mit dem Konstrukt *Persönlichkeit* bleibt jedoch unberücksichtigt. Ebenso

stellt Hattie in seiner Studie (2013) zwar fest, dass letztlich die Lehrkraft es ist, die die Unterrichtsgestaltung bestimmt und daher ausschlaggebend für gelungenen Unterricht ist: So gelten ihm zufolge die Lehrerinnen und Lehrer als entscheidende Faktoren für guten Unterricht resp. Lernerfolg seitens der Schülerinnen und Schüler im Sinne messbarer kognitiver Fachleistungen. Da zudem äußere Faktoren wie Schul- und Unterrichtsstrukturen in seinem metaanalytischen Ranking relativ weit unten in ihrer Wirkkraft auf Schülerleistungen stehen, schließt er, dass Unterschiede in der Qualität nicht auf Schulebene, sondern auf Ebene der einzelnen Klassen resp. der einzelnen Lehrkräfte und deren Person zu finden sind². Doch auch bei Hattie fehlt jegliche Ausdifferenzierung, was nun deren Persönlichkeit genau sein soll – abgesehen von der Forderung, dass die Lehrpersonen die Schülerperspektive einnehmen sollen, indem sie sich selbst permanent reflektieren und zudem Schülerfeedbacks/-statements einholen (vgl. Hattie 2013). Obwohl die oft zitierten Studien keine entsprechenden oder eindeutigen Rückschlüsse diesbezüglich zulassen, werden derartige Äußerungen (u.a. auch von Kunter (2013), die sich jedoch explizit auf den personalen Aspekt des Enthusiasmus³ bezieht, der in einer enger gefassten und systematischen Konstruktion von Persönlichkeit jedoch nicht als Persönlichkeitsmerkmal gilt) zur theoretischen Grundlage bzw. als Nachweise für Zusammenhänge von Lehrerpersönlichkeit und Unterrichtsqualität herangezogen. Es lesen sich nicht selten Schlussfolgerungen ähnlich dieser:

„Teachers have a crucial impact on students' achievement (Hattie, 2009). [...] Individual characteristics, such as [...] personality, have been assumed to have both direct and indirect effects on teacher quality and success (Kennedy et al., 2008; Kunter et al., 2013) in terms of students' learning (Hattie, 2009, 2012).“ (Hench/Klusmann/Lüdtke/Trautwein 2015, S. 46/53)

In der Auseinandersetzung mit dem Thema *Lehrerpersönlichkeit* findet sich lediglich bei Mayr (2003, 2014 mit Neuweg 2006, 2018) eine klare Vorstellung: Hier steht die Selbsteinschätzung der Persönlichkeit von (angehenden) Lehrkräften im Vordergrund – allerdings nicht in direktem Zusammenhang mit der Unterrichtsqualität, sondern unter dem eher gesundheitlich-ökonomischen Aspekt

² Wohingegen Klieme (2013) die Befunde als Beleg für die Bedeutung von Klassenführungs-kompetenzen im Sinne der Strukturiertheit des Unterrichts und im weiteren Sinne als Beleg für die Bedeutung von Prozessmerkmalen des Unterrichts interpretiert.

der Berufsbelastungstauglichkeit sowie dem Studienerfolg. Es gilt dabei die eigene Eignung und Motivation für den Lehrerberuf selbst zu erkunden. Ein populäres Instrument hierfür ist etwa der CCT (Career Counselling for Teachers), ein von Johannes Mayr in Österreich entwickeltes Online-Verfahren (vgl. dazu Mayr 2009), welches außerdem in Nordrhein-Westfalen und Baden-Württemberg als Selbsterkundungsverfahren eingesetzt wird. Dieses Instrument beschäftigt sich mit sogenannten Persönlichkeitskonstrukten wie Extrovertiertheit und Introvertiertheit, die als nicht oder kaum veränderbar gelten. Hiermit ist die Schwierigkeit verbunden, dass eine Rückmeldung auf Persönlichkeitskonstrukte durch einen Test kaum Entwicklungspotenzial bietet, denn stabile Aspekte der Persönlichkeit entwickeln sich nur geringfügig in die eine oder andere Richtung. Andere, nicht an stabilen Persönlichkeitskonstrukten orientierte Verfahren, die sich eher an veränderbaren Verhaltensmustern orientieren wie z. B. AVEM (Arbeitsbezogene Verhaltens- und Erlebensmuster (siehe Schaarschmidt/Fischer 2008)), konzentrieren sich auf das Erleben und die Verarbeitung von Belastungen.

So wurde Persönlichkeit lange als eine *unantastbare* Größe angesehen, als manifester wie auch intimer Bereich. Allerdings stellt Persönlichkeit ebenso eine *äußere* Größe (nicht zu verwechseln mit einer vermeintlich objektiven) dar: als die in der äußeren Welt in Erscheinung tretenden Anteile des Subjekts oder einer Person *an sich*. Innen- und Außenwelt fungieren dabei als Dimensionen eines fluiden Wechselspiels.

Vor allem Clausen (2002) hat mit seiner Untersuchung zur Perspektivenabhängigkeit von Wahrnehmungen der Unterrichtsqualität die Gültigkeit postmoderner Ideen (wie Multiperspektivität, Diversität, gleiche Gültigkeit unterschiedlicher Realitäten nebeneinander) für die Schul- bzw. Unterrichtsforschung empirisch belegt. Mit Clausen gelten verschiedene Perspektiven als perspektivenabhängig valide, wonach es nicht zweckmäßig erscheint, Lehrer- und Schülervariablen/-wahrnehmungen zusammenzubringen. Allerdings liegen keine systematischen, der Vergleichbarkeit dienlichen Untersuchungen zum Zusammenhang von Lehrerpersönlichkeit und Unterrichtsqualität aus Schülersicht vor. Es gilt nun also, bestehende Forschungsbefunde dem aktuellen Stand der Perspektiven-spezifität anzupassen und entsprechend zu aktualisieren.

Für das Konstrukt der *Lehrerpersönlichkeit* liegt diesbezüglich insgesamt ein sogenannter doppelter Forschungsstand vor. Das bedeutet, dass die theoretischen Befunde anders als die empirischen Befunde gelagert sind. Während in der Theorie *Persönlichkeit* vornehmlich entlang postmoderner Perspektiven gedacht

wird, basiert die empirische Forschung zur *Lehrerpersönlichkeit* in weiten Teilen auf Perspektiven moderner Theorien neuhumanistischer Prägung (vgl. Bohnsack 2004). Konkret zeigt sich dieser Umstand darin, dass neuere Subjekttheorien gedankliche Konstrukte von *Persönlichkeit* mehrdimensional als komplex-wechselwirksame Modelle³ innerhalb sozialer Interaktionen einordnen (siehe hierzu bspw. Meyer-Drawe 2000, Fromme 1997), was der in der empirischen Schul- und Unterrichtsforschung gängigen, eindimensionalen Methode der Selbstbefragung zur Erfassung der Lehrerpersönlichkeit gegenübersteht.

Diese Diskrepanz innerhalb der Forschung zur *Lehrerpersönlichkeit* zieht sich bis heute nahezu nahtlos durch die Schul- und Unterrichtsforschung. Während es innerhalb der Schulforschung allerdings neue Ansätze gibt, die die Mehrdimensionalität von Untersuchungsgegenständen wie der Unterrichtsqualität abhängig von der jeweiligen Wahrnehmungsperspektive belegen (siehe bspw. Clausen 2002, Gruehn 2000), scheint diese Erkenntnis für die Wahrnehmung der Lehrerpersönlichkeit nicht zu gelten, sondern diese wird empirisch weiterhin behandelt, als sei sie eine identische Einheit innerer und äußerer Aspekte, die mittels reflektierender Selbstbefragung als *ein Ding an sich* erfasst werden könnte, das in keinem Gegensatz zu den Schülerwahrnehmungen steht.

1.1 Perspektivität und Wahrnehmung

„Das Ausmaß der subjektiven Identifikation eines Individuums mit einer sozialen Gruppe oder Institution beeinflusst seine Urteile, Meinungen, Entscheidungen und auch schlussendlich seine Handlungen. Diese These, dass Identitäten aufgrund eines Darstellungsbedürfnisses motivational auf das Verhalten wirken, bildet auch eine Grundannahme des strukturellen symbolischen Interaktionismus. Er stellt die These auf, dass man nicht nur Erwartungen das Verhalten anderer Personen betreffend hegt, sondern dass man auch sich selbst klassifiziert, was zu verinnerlichten Erwartungen an das eigene Handeln führt. [...] Nun verfügt aber jeder Mensch über mehrere soziale Identitäten (Familie, Beruf, Freundeskreis, Gemeinde, Staat, etc.).“ (Haller/Müller 2006, S. 16)

³ Innerhalb der Sprachspieltheorie beziehen sich diese bspw. auf das Konzept des Ausdrucks (siehe vertiefend hierzu Schäfer 2005) und das Konzept der narrativen Identität (siehe vertiefend hierzu Paetow 2004).

Diesem Zitat folgend formulieren Back und Vazire (2012) die Übereinstimmung des Selbstbildes einer Person mit der realen Persönlichkeit als eine naive Definition von *Selbsterkenntnis*. Präziser ließe sich demgegenüber Selbsterkenntnis als akkurate explizite Selbstwahrnehmung bzgl. der eigenen Gedanken, Gefühle und Handlungen und ein Bewusstsein für die Interpretationen dieser durch andere, also als Übereinstimmung des Selbstbildes mit Verhaltensbeobachtungen und Reputationen, ausmachen. In einem derartigen Prozessmodell der Persönlichkeitsselbsterkenntnis (also wie Persönlichkeit im sozialen Kontext sowohl ausgedrückt als auch wahrgenommen wird) sind oder wären Veränderungen des Selbstbildes demnach durch die Integration der Reputationen anderer in dieses Selbstbild möglich, aber:

„People are very resistant to changing their self-concept even when facing relatively clear-cut information and contradictory information. [...] Not only did people resist changing their perceptions of their own personality, they also failed to change their perceptions of how they come across to others and how they behave.“ (Back/Vazire 2012, S. 148)⁴

Es besteht also eine gewisse Konsistenz des Selbstbildes aufgrund von resistenzbildenden, psychologischen Prozessen wie Selbstüberschätzung (sichtbar anhand selbstdienlicher Bias⁵) oder Selbstbestätigung (fundamentales Bedürfnis nach Bestätigung des bestehenden Selbstbildes). Resistenzen gegen Fremdeinschätzungen basieren vor allem auf einem stabilen Selbstkonzept, dessen höchste Ausprägung bei gesunden Erwachsenen zu finden ist (Srivastava 2012). Andererseits entspricht die Intention, etwas (non-verbal) auszudrücken, nicht immer dem, was tatsächlich ausgedrückt wird, d.h. es gibt Beschränkungen bei der Übersetzung von Intentionen in den entsprechenden Ausdruck: (non-verbale) Verhaltensweisen im Sinne von (weitgehend) spontanen, nicht regulierten Äußerungen innerer Vorgänge⁶ fungieren dabei als Bias von intendiertem Verhalten (De Paulo 1992).

⁴ Diese Feststellung bezieht sich auf eine Studie von Hofmann/Gschwendner/Schmitt (2009).

⁵ Bias sind Variablen oder Faktoren, die das Antwortverhalten beeinflussen und so zu verzerrten Ergebnissen oder Interpretationen führen können. Bias wird aus dem Englischen dementsprechend auch mit dem Begriff Verzerrung übersetzt (Wirtz 2014).

⁶ Im Sinne der Theorien von bspw. Allport (1961) oder Goldstein (2010), die im Gegensatz zu ethologischen und behavioristischen Perspektiven stehen.

Hinzu kommen die eigenen blinden Flecken im Selbstbild, d.h. explizites und implizites Selbstbild stechen als *nichtidentisch* heraus (vgl. Back/Vazire 2012). Dieser Ansatz stellt eine Übertragung der Theorie des expliziten und impliziten Wissens auf das Selbst dar. Demzufolge bildet das explizite Selbst den Anteil der Person, der sich bewusst und reflektiert auszudrücken vermag, während das implizite Selbst unbewusst und unreflektiert in Erscheinung tritt und somit die blinden Flecken im Selbstbild verursacht (vgl. Polanyi 1985, zur vertieften Auseinandersetzung siehe Nonaka/Takeuchi 1997). Viele Befindlichkeiten und Verhaltensweisen laufen automatisiert ab und sind für unser Bewusstsein nicht direkt erfassbar. Das Bewusstsein scheitert zudem häufig an der Integration von impliziten Aspekten der Persönlichkeit. D.h. unterbewusste, vor allem nonverbale Verhaltensweisen stehen außerhalb der expliziten, bewussten Wahrnehmung und finden schwerlich den Weg ins Bewusstsein. Sie sind zwar generell, aber nur teilweise regulierbar (De Paulo 1992). Das Selbstbild kann durch Konfrontation mit Fremdeinschätzungen also bedingt modifiziert werden (Srivastava 2012):

„As a case in point, people seem to be not very willing to integrate [...] self-perceptions of their nonverbal behavior (most of which is impulsively) into their explicit self-concept.” (Back/Vazire 2012, S.143)⁷

Blinde Flecken hinsichtlich nonverbaler Verhaltensweisen des Unterbewusstseins bleiben allerdings trotz einschlägiger Hinweise auf sie bestehen; lediglich die reflektierten, kontrollierten Anteile des sichtbaren Verhaltens werden vom expliziten Selbstkonzept erfasst. Es zeigen sich also introspektive Grenzen: Das Bewusstsein verfügt immer nur über einen eingeschränkten Zugriff auf bestimmte Aspekte des Selbst, die das alltäglich an den Tag gelegte Verhalten beeinflussen (Back/Vazire 2012).

Selbsteinschätzungen haben zwar den Vorteil der Zugänglichkeit zum Wissen über die inneren Vorgänge. Daher werden sie vornehmlich zum Erheben und Messen von Persönlichkeitsmerkmalen in der Forschung gewählt, obgleich die Verzerrungsgefahr durch introspektive Begrenzungen, Selbstverstärkung/-aufwertung sowie sozial erwünschtes Antwortverhalten bei dieser Perspektive hoch ist. Während bei anderen die Verzerrung des Selbstbildes erkennbar ist und Selbsteinschätzungen skeptisch betrachtet werden, schließen sich in der Regel

⁷ Nachgewiesen wurde dieser Befund in einer Studie von Hofmann/Gschwendner/Schmitt (2009).

die meisten Menschen selbst von dieser Erkenntnis aus. Die Verknüpfungen zwischen expliziten Selbstbildern und impliziten Persönlichkeitsmerkmalen, Verhaltensweisen und Reputationen unterliegen ähnlichen Bias: So stellen bspw. Selbstaufwertungsprozesse, Konsistenzprozesse (Eigenschaften und Reputationen werden als kohärente, d.h. zeit- und situationsübergreifende Einheit wahrgenommen), Vergleichseffekte (eigene Merkmale werden mit denen anderer verglichen) Biasfaktoren⁸ dar. Zur Steigerung der Übereinstimmung von implizitem und explizitem Selbstbild rät Funder:

„Travel. Meet the kind of people you do not ordinary meet. Most important, be sure to seek feedback.” (Funder 1999, S. 205)

Dennoch sind Selbsteinschätzungen auch nicht bloß Illusionen des Selbst, sondern tauchen durchaus unterschiedlich ähnlich auch in Beobachtungsberichten auf. An dieser Stelle soll jedoch herausgestellt werden, dass das Selbstbild nicht derart klar, überlegen und unverfälscht ist wie oftmals und vor allem von uns jeweils selbst angenommen wird, woraus folgt, dass das Selbstbild keinesfalls die *wirkliche* Persönlichkeit darstellt (Back/Vazire 2012). Was also ist dann die *wirkliche* Persönlichkeit? Nach Back und Vazire (2012) ist diese als solche nicht definierbar und auch nicht empirisch messbar, aber dennoch auf verschiedene begründbare Arten und Weisen erfassbar als explizites Selbstbild (Selbstbericht) oder implizites Selbstkonzept (Metabericht), durch Verhaltensbeobachtungen oder Reputationen (Fremdberichte) (mehr dazu in Kap. 3).

Zwar ist die häufigste Methode zur Persönlichkeitserfassung anhand von Merkmalen wie bspw. der BigFive (vgl. Kap. 3.2) noch immer die Selbsteinschätzung resp. der Selbstbericht: in dieser Form stellt sich *Persönlichkeit* als die Reflexion der eigenen Identität und somit als eine intra-psychische Repräsentation des Selbst dar (Roberts/Wood 2006). Da sich diese Persönlichkeitsmerkmale jedoch auch in der Umwelt des Individuums niederschlagen, sind solche Merkmale auch von außen sowohl beobachtbar als auch zuschreibbar. Demnach stellt die Fremdeinschätzung der Persönlichkeit eine weitere mögliche und gleichermaßen gültige Methode zur Erhebung von Persönlichkeitsmerkmalen dar (bspw. vergleichbar mit dem Ruf einer Person) (Schaffhuser 2014). In sozialen Interaktionen, die

⁸ Die Fähigkeit, der Selbstbeschreibung entsprechend zu handeln und damit auch die Übereinstimmung von Bewusstsein und tatsächlichem Verhalten zu generieren, gilt im Übrigen ebenfalls als von derartigen Bias beeinträchtigt (vgl. Back/Vazire 2012).

hier als symmetrischer Prozess gelten, sind Personen immer gleichzeitig sowohl Akteure als auch Beobachter (Wirtz 2014). Das bedeutet zunächst, dass zwischen diesen beiden Perspektiven unterschieden werden muss (Abele 2014) und darüber hinaus, dass die Wahrnehmung innerhalb sozialer Interaktionen einem ständigen Wechsel der Perspektiven unterliegt bzw. dieses Wechsels bedarf (Wirtz 2014). Der auf dieser Erkenntnis basierende Interaktionismus⁹ ist eine Reaktion auf die situationistische Kritik am Eigenschaftsmodell der Persönlichkeit und

„[...] besagt, dass Merkmale der Person den Einfluss der Situation und Merkmale der Situation den Einfluss der Persönlichkeit einer Person auf deren Erleben und Verhalten moderieren.“ (Wirtz 2014, S.803)

Diesem Ansatz folgt auch ein in das Projekt „Wege in den Beruf“ (Neuber/Künsting/Lipowsky 2014) eingebetteter Fokus, der auf dem gezeigten Verhalten von Lehrpersonen bzgl. des Enthusiasmus⁴ liegt. Der Enthusiasmus wird dabei als kontextueller Stimulus eingeordnet, der das Lernen der Schüler beeinflusst. In Person-Situation-Interaktionen ist daher sowohl die Messung relevanter Persönlichkeitsmerkmale als auch relevanter Situationsmerkmale erforderlich, da Menschen bspw. identische Situationen je nach Persönlichkeit unterschiedlich wahrnehmen. Diese sogenannten reaktiven Unterschiede stehen neben Unterschieden evokativer (Personen beeinflussen und reagieren unterschiedlich auf andere Person), proaktiver (Suche nach bestimmter Art von Einflüssen) und manipulativer (spezifische Gestaltung von Situationen) Art (Wirtz 2014). Unter anderem verweist dies auf die Kontextspezifität von Persönlichkeitswahrnehmungen (sowohl in der Selbst- als auch der Fremdperspektive).

Zwischenmenschliche Beziehungen basieren auf einer dynamischen Wechselwirkung zweier Persönlichkeiten in Form weitgehend stabiler Interaktionsmuster, Interaktionsskripte oder Interaktionsstile, die unter den beteiligten Persönlichkeiten ausgehandelt werden müssen: Die Psychologie der Beziehung geht damit über die Psychologie der Persönlichkeit hinaus (Asendorpf/Neyer 2012).

⁹ Dieser gründet auf dem symbolischen Interaktionismus von Mead (1934), wonach das, was andere über uns denken in deren Verhalten uns gegenüber reflektiert wird. Dieses nehmen wir wiederum wahr, um uns zu evaluieren (sogenannte Feedback-Integration) (vgl. Back/Vazire 2012).

„Beurteilungen durch Interaktionspartner (statt durch Beobachter) sind nur dann sinnvoll, wenn es um den subjektiven Eindruck von der Versuchsperson speziell bei Interaktionspartnern geht [...]“ (Asendorpf/Neyer 2012, S. 101)

Grundlage dafür ist bestenfalls die alltägliche Selbstpräsentation in Form des Verhaltens, das eine Person *an den Tag legt*. Diese Bild basiert dann auf vielen Fallbeobachtungen und Anekdoten (De Paulo 1992), selbstevaluative Merkmale können dabei nicht erhoben werden. So gehört bspw. das Selbstwertgefühl zur Familie der selbstevaluativen Persönlichkeitsmerkmale, da es per definitionem die individuellen Gefühle und Bewertungen der eigenen Person zu sich selbst als Person betrifft (Rosenberg 1995).

Bzgl. der Metapräzision – womit die eigene Einschätzung der Sichtweisen anderer (Fremdbilder) gemeint ist; als Vorstellung von Individuen davon, wie andere sie sehen – zeigen sich durchwachsene Befunde: erfolgt sie in Bezug auf jemand bestimmten, ist sie schwach; die Einschätzung von anderen in bestimmten Kontexten jedoch ist stärker bzw. treffsicherer (Back/Vazire 2012). Dieser Befund stärkt v.a. die Annahme einer umfassenden Kontextspezifität des persönlichen Auftretens (vgl. Kap. 3.4.2). Insgesamt zeigt sich faktorenanalytisch, dass alle drei Wahrnehmungsperspektiven – Selbsteinschätzung, Beurteilungen durch Interaktionspartner, Metapräzision – Übereinstimmungen, aber ebenso einzigartige Aspekte aufweisen: Sie hängen zusammen, aber sind dennoch verschieden voneinander (Back/Vazire 2012).

Für die Schule ist die Frage nach der Perspektivität und der damit einhergehenden Wahrnehmung ebenfalls von besonderer Bedeutung, da Schülerurteile, Schülerfeedbacks, Schülereinschätzungen etc. vermehrt in den Fokus geraten. So zeigen sich bei Fremdberichten im Rahmen von Schule relativ hohe Übereinstimmungen von Schülerinnen und Schülern innerhalb einer Schulklasse, deren Wahrnehmungsperspektive vermehrt eine eigenständige Validität zugesprochen wird: und zwar als geteilte/kollektive Wahrnehmung, dem sogenannten Klassenklima, anstatt als Messfehler. Heute stellt sich daher eher die Frage nach der richtigen Perspektive (Schüler-/Beobachter-/Lehrersicht) für die jeweilige Fragestellung oder das jeweilige Instrument, wobei jede Perspektive ihre eigenen Stärken und Schwächen hat (Gruehn 2000, Wagner 2008, Clausen 2002). Für Schülerurteile z.B. sind bei der Itemformulierung die Wahrnehmungsperspektive (individuell oder aus Sicht der Klasse), der Inferenzgrad (niedrig=konkret, hoch=abstrakt) und die Adressaten des Lehrerverhaltens (Klasse oder Subjekt) relevante Faktoren (vgl. Wagner 2008). Personenmerkmale – sowohl von Lehrkraft als auch von

deren Schülerinnen und Schülern – gelten in diesem Zusammenhang als möglicherweise bedeutsame Bias (im Sinne eines Halo-Effekts) (Wagner 2008, Gruehn 2000). Gruehn (2000) kritisiert in diesem Zusammenhang die verschiedenen Ansätze zur strukturierten Erhebung der Unterrichtsqualität dahingehend, dass sie unsystematisch mit Begrifflichkeiten umgehen und dabei unterschiedliche Konstrukte fast gleich benennen. Betreffend die Lehrerpersönlichkeit kann sich dem nur angeschlossen werden.

Helmke (u.a. 2009, 2014) weist mit seinem Angebots-Nutzungs-Modell (der Lehr-Lern-Forschung) ebenfalls auf eine von den Wahrnehmungen der Schülerinnen und Schülern abhängige Kausalkette hin, die zwischen den Prozessmerkmalen des Unterrichts und seinen Produkten (wie etwa dem Lernerfolg oder der Zufriedenheit etc.) liegt. In diesem Zusammenhang kommt die Frage auf, wann (non-verbales) Verhalten willentlich oder nichtwillentlich hervorgebracht werden kann oder hervorgebracht wird (De Paulo 1992), um bspw. auch über solche Aspekte Einfluss auf die mögliche Wahrnehmung und somit die Unterrichtsprodukte zu nehmen. Es zeigte sich, dass dahingehend Motivation und Zuversicht durchaus mögliche Einflussfaktoren auf die Verhaltenssteuerung sind; es gibt demnach Bedingungen, unter denen es Menschen gelingt, ihr Verhalten für den Zweck der Selbstdarstellung gezielt zu regulieren¹⁰ (De Paulo 1992). So gesehen, kann das persönliche Auftreten der Lehrkraft in Form der bewusst regulierten Selbstdarstellung durchaus als Teil der beruflichen Kompetenz aufgefasst werden.

„Das Kompetenzverständnis orientiert sich am Begriff der „beruflichen Kompetenz“ Terharts, in dem wiederum der von Weinert definierte Kompetenzbegriff durch einen verstärkten Blick auf die Performanz bzw. auf die Handlung in einer bestimmten Situation erweitert [...] wird.“ (Unterweger 2014, S.2)

¹⁰ Der Grad an Übereinstimmung von implizitem und explizitem Selbstbild wird anhand des Zusammenspiels von reflektierten und impulsiven Kognitions-, Gefühls- und Handlungsprozessen in ihrer jeweils individuellen, typischen Art interpretiert. Durch Veränderungen von Bewusstseinszuständen wird ein Zugang zu verändertem Selbstwissen ermöglicht, das sich dann wiederum in einer veränderten Selbstrepräsentation äußert (z. B. durch Schreiben über sich bzw. Elaboration oder intentionale Überarbeitung expliziter Persönlichkeitsaspekte wird die Übereinstimmung von implizitem und explizitem Selbstbild gesteigert) (Back/Vazire 2012).

Im Weiteren stellt nach Unterweger (2014)

„[d]er Kompetenzansatz [...] die Lernenden ins Zentrum des Bildungsprozesses [stellt]; die Dreiteilung in Selbst-, Sozial- und Systemkompetenz bedeutet daher für die Professionalisierung von Pädagog/inn/en insofern eine adäquate Struktur für die Persönlichkeitsarbeit, als der individuelle, sich bewusst wahrnehmende und reflektierende Mensch gleichzeitig als agierendes, erlebendes Subjekt und als Objekt (der Wahrnehmung und Reflexion) angesprochen ist:

- im Kontakt und in Beziehung zu sich selbst (Bereich der Selbstkompetenz),
- im Kontakt und in Beziehung zur jeweiligen Bezugsgruppe (Bereich der Sozialkompetenz) und
- im Kontakt und in Beziehung zur jeweiligen Institution, Organisation, Gesellschaft und kulturellen Gemeinschaft (Bereich der Systemkompetenz).“ (Unterweger 2014, S. 3)

Abschließend wird zusammengefasst, dass sowohl in Bezug auf die *Persönlichkeit* als auch die *Unterrichtsqualität* die einer Wahrnehmung zugrunde liegende Perspektivität von essentieller Bedeutung für die Anlage und die Interpretation von Befunden ist.

1.2 Aufbau der Arbeit

Nach dieser ersten Einführung in die grundlegende theoretische Ausrichtung der vorliegenden Arbeit – nämlich der Abhängigkeit subjektiver Wahrnehmung von der jeweiligen Perspektive – werden im folgenden Theorieteil (I.) sowohl die Konzepte und ihre Operationalisierungen als auch aktuelle und relevante Forschungsbefunde hierzu dargestellt. Dazu wird zunächst das Konstrukt der *guten* Lehrkraft theoretisch sowie anhand seines aktuellen Forschungsstands vorgestellt (Kap. 2). Da das Konstrukt der *guten* Lehrkraft als Synthese von Persönlichkeit und Unterrichtsqualität aufgefasst wird, greift der Forschungsüberblick zur *guten* Lehrerpersönlichkeit (Kap. 2.2) etablierten und für den Verlauf der Arbeit relevanten Konstrukten von Persönlichkeit und Unterrichtsqualität vor. Danach erst wird die *gute* Lehrkraft systematisch in seine Einzelkonstrukte – Persönlichkeit (Kap. 3) und Unterrichtsqualität (Kap. 4) – aufgespalten und präzi-

siert. Die relevanten Theorien und Diskurse jener gedanklichen Konstrukte werden dabei erläutert und einschlägige Forschungsbefunde zusammengefasst. Die jeweiligen Konstrukte werden zunächst definiert, dann von anderen, verwandten Konstrukten abgegrenzt und dadurch begrifflich eingegrenzt. Dabei werden jeweils die theoretischen Grundlagen vorgestellt, die einen Überblick über die Betrachtungsweisen der Konstrukte liefern. Auf diesem Weg wird ein Arbeitsbegriff herausgefiltert, der das Verständnis der untersuchten Gegenstände innerhalb dieser Arbeit klärt (in Kap. 2 jedoch bereits eingebettet ist). In den Kapiteln zu Persönlichkeit und Unterrichtsqualität werden daran anschließend die Orientierungen und Paradigmen der jeweiligen Forschungsfelder nachgezeichnet (Kap. 3.1, 4.1 – in Bezug auf die Lehrperson bilden sich die entsprechenden Orientierungen und Paradigmen aus jenen der Persönlichkeit und der Unterrichtsqualität heraus), um im weiteren Verlauf der Arbeit das eigene Forschungsverständnis darin zu verorten. Im Anschluss erfolgt – erneut je Konstrukt – die Auseinandersetzung mit deren empirischen Erfassbarkeit resp. mit Möglichkeiten zur Operationalisierung von *Persönlichkeit*, *Unterrichtsqualität* und ihrer Implikation im Konstrukt *Lehrerpersönlichkeit* (Lehrperson s. Kap. 2.1, Persönlichkeit s. Kap. 3.2, Unterrichtsqualität s. Kap. 4.2), woraufhin vor allem hinsichtlich der Konstrukte *Persönlichkeit* und *Unterrichtsqualität* jeweils eine valide und erprobte Operationalisierung begründet ausgewählt und vertieft in ihrer Entwicklung und Systematik (Kap. 3.3, 4.3) erläutert wird. Daraufhin wird für jedes der verschiedenen Forschungsfelder der aktuelle Forschungsstand aufbereitet (Kap. 2.2 und 2.3, 3.4, 4.4). Den Theorieteil schließen die auf Basis der theoretischen Überlegungen und Befunde abgeleiteten Fragestellungen dieser Arbeit ab (Kap. 5).

Im Methodenteil (II.) werden zunächst das zur Untersuchung der in Kap. 5 erarbeiteten Fragestellungen erforderliche Untersuchungsdesign (Kap. 6.1) und die Stichprobe der erhobenen Daten (Kap. 6.2) beschrieben. Dem folgen theoretische Abwägungen zum Umgang mit der Mehrebenenstruktur der in Schulklassen gebündelten Daten (Kap. 7), zu statistischen Möglichkeiten im Umgang mit fehlenden Angaben (Kap. 8) und zur Interpretation der Güte statistischer Modelle (Kap. 9). Nach einem Überblick über die Grundlagen von Erhebungsinventar und Validierungsverfahren (Kap. 10) werden die einzelnen Instrumente zur Erhebung der Konstrukte (*Lehrer-*)*Persönlichkeit* (Kap. 11) und *Unterrichtsqualität* (Kap. 12) präzise beschrieben und dabei inhaltlich validiert (11.2, 12.2) sowie auf ihre statistische Reliabilität und Validität geprüft (Kap. 11.3, 12.3). Da das

zur Erhebung der Lehrerpersönlichkeit aus Schülersicht eingesetzte Instrument bislang nicht hinsichtlich Fremdeinschätzungen erprobt ist, fällt dessen Validierung recht umfangreich aus. Der methodische Teil wird mit Erläuterungen zu den zur Beantwortung der Fragestellungen dieser Arbeit notwendigen Analysestrategien abgeschlossen (Kap. 13).

Im Ergebnisteil (III.) werden zuerst die deskriptiven Befunde (Kap. 14) zur Selbsteinschätzung der Lehrerpersönlichkeit (Kap. 14.1), die durch die Schülerinnen und Schüler eingeschätzte Lehrerpersönlichkeit (Kap. 14.2) und die Unterrichtsqualität aus Schülersicht (Kap. 14.3) vorgestellt. Die Werte der schülerperzipierten Lehrerpersönlichkeit und Unterrichtsqualität werden dabei – als Folge ihrer Validierung (s.o.) – fachspezifisch für die Fächer Deutsch (Kap. 14.2.1, 14.3.1) und Mathematik (Kap. 14.2.2, 14.3.2) berichtet. Danach werden die korrelativen Befunde zwischen den durch die Schülerinnen und Schüler fremdeingeschätzten und den durch die Lehrerinnen und Lehrer selbsteingeschätzten Lehrerpersönlichkeit der ersten Forschungsfrage dargestellt (Kap. 15). Darauf folgend werden die Ergebnisse der Regressionsanalysen zwischen schülerperzipierter Lehrerpersönlichkeit und Unterrichtsqualität (Kap. 16) abgebildet.

Im letzten Teil (IV. Diskussion) werden zunächst die einzelnen Befunde interpretiert (Kap. 17) und anschließend zusammengefasst u.a. kritisch diskutiert (Kap. 18), woraus sich zum einen ein Fazit bzgl. möglicher praktischer Verwendungen der empirischen Ergebnisse dieser Arbeit (Kap. 19) sowie anhand der Limitationen der durchgeführten Untersuchung Ausblicke für weitere Forschungen ableiten lassen (Kap. 20).

I. Theorien

2 Die *gute* Lehrkraft als Synthese von Persönlichkeit und Unterrichtsqualität

Mindestens seit den 1920ern beruhte die zu dieser Zeit noch rein geisteswissenschaftliche Lehrerforschung ausschließlich auf dem Persönlichkeitsparadigma (Mayr/Neuweg 2006). Diesem Paradigma liegt vornehmlich eine neuhumanistische Ausrichtung zugrunde, die sich im Spannungsfeld zwischen Vorstellungen eines mittels Bildung geförderten, kontinuierlichen und harmonischen Fortschreitens zum Ziel der Vervollkommnung des Selbst und Vorstellungen autonomer Anteile von Individuen verorten lässt (Bohnsack 2004, siehe hierzu vertiefend Meyer-Drawe 2000). Ebenso wie auch reformpädagogische Ansätze hatte diese Ausrichtung hinsichtlich der Lehrerforschung zur Folge, dass entlang kultur-historisch wandelbarer Normen Idealbilder *guter* Lehrer und Lehrerinnen entworfen wurden (Bohnsack 2004). Vor allem die Praktiker pädagogischer Tätigkeitsfelder heben immer wieder hervor, wie bedeutsam die Persönlichkeit im Zusammenhang mit ihrem Beruf ist. Wohingegen der wissenschaftliche Ansatz immer wieder von derartigen, nicht erlernbaren und demnach nicht beeinflussbaren Voraussetzungen Abstand zu nehmen versucht (Mayr/Neuweg 2006).

Ebenso wie in Bezug auf die Unterrichtsqualität wurden Vorstellungen des *guten* Lehrers bzw. der *guten* Lehrerin bis in die 1960er/1970er Jahre also theoretisch und vor allem ideell geprägt (Terhart 2012). Mit der Konsequenz, dass bis heute der Begriff der Lehrerpersönlichkeit stark normativ belastet ist (Mayr/Neuweg 2006, Bohnsack 2004).

Erst mit der empirischen Wende Ende der 1960er, Anfang der 1970er Jahre wurde der Aspekt des Einflusses der Lehrerpersönlichkeit auf Lehr-Lernprozesse wieder aufgegriffen. Dabei wurde er als unabhängiger Wirkfaktor (Bedingungsvariable) in Bezug auf Schülerleistungen betrachtet (Haag/Lohrmann 2009), was allerdings zunächst als eine Vermischung normativer und empirischer Aussagen entlang normativer Tugendkataloge eher trivial sowie messverfahrenstechnisch rückständig blieb (vgl. Mayr/Neuweg 2006). Angesichts des Problems der Normativität und der damit verbundenen historischen Belastung des Persönlichkeitsbegriffs erfolgte während der 1970er/1980er Jahre nahezu keine wissenschaftliche Auseinandersetzung mit der Persönlichkeit des Lehrers resp. der Lehrerin, sondern es wurden vor allem Schülermerkmale innerhalb des Prozess-Produkt-Paradigmas untersucht (Bohnsack 2004, Terhart 2012). Die meisten danach sehr

bald wieder einsetzenden empirischen Forschungen zur Person der Lehrerinnen und Lehrer basieren dabei ebenfalls auf dem Prozess-Produkt-Paradigma. Diesem Paradigma zufolge besteht ein Zusammenhang zwischen der persönlichen Art des Unterrichtens seitens der Lehrkraft und den entsprechenden Lerneffekten seitens der Schülerinnen und Schüler (Terhart 2007). Da sich die empirische Forschung schwer damit tut, ein ideales, normativ ausgerichtetes Lehrerbild zu entwerfen (Terhart 2007), konzentriert sie sich seit den späten 1980er Jahren in erster Linie darauf zu beobachten, was *tatsächlich ist* (Asendorpf 2004), statt darauf, was vermeintlich *sein sollte* (Haag/Lohrmann 2009). Dennoch wird der Lehrerforschung weiterhin vorgeworfen, gemäß dem Vorgehen der traditionellen, im Ursprung bereits in der Antike wurzelnden Lehrerforschung Idealbilder von Lehrerpersönlichkeiten zu entwerfen, dem angehende Lehrerinnen und Lehrer dann entsprechen sollen (Haag/Lohrmann 2009).

In den 1990er Jahren entwickelte sich innerhalb der Lehrerforschung die *Professionalität* als Zielperspektive. Diese war angelehnt an das eher simplifizierte Expertenparadigma, innerhalb dessen gemäß der kognitiven Wende (mit dem Fokus auf dem Denken von Lehrerinnen und Lehrern) davon ausgegangen wurde, dass sich Kompetenz allein durch Wissen und Können herausbildet. Lehrerhandeln wird dann als das Resultat von Expertenwissen betrachtet (Bromme 1997). Allerdings scheint dieser Ansatz zu optimistisch oder gar illusionär, zumal entsprechende empirische Belege für die Wirksamkeit von Lehrerbildung sowie Lehrerfortbildungen widersprüchlich sind (Mayr/Neuweg 2006). Als Experten gelten zudem (normativ entlang der Schülerleistungen definiert) erfolgreiche Lehrkräfte, die gewisse Kennzeichen in Unterrichtssituationen aufweisen (z.B. eine kategoriale Einordnung von Unterrichtsgeschehen vornehmen und/oder flexibel auf Unterrichtssituationen reagieren, ohne ihr angestrebtes Ziel aus den Augen zu verlieren) und über bestimmte Wissensbereiche verfügen. Als gelingendes Beziehungsgeflecht zwischen Lehrerkognitionen und Lehrerhandeln „muss das professionsbezogene Wissen immer in die personale Entwicklung einer Lehrkraft integriert sein [...]“ (Haag/Lohrmann 2009, S. 466). Dieses an einem derartigen Expertenparadigma orientierte Verständnis von Professionalität entwickelte sich bald weiter zu einem kompetenzorientierten Modell (vgl. Mayr/Neuweg 2006, Terhart 2012). Dabei zeichneten sich verschiedene Konzepte zur Professionalisierung im Lehrerberuf ab, die sich in zwei Strömungen unterteilen lassen: solche, die traditionell normativen Vorstellungen entsprechen, und solche, die sich entlang des *New Professionalism* einordnen lassen. *Traditionell*

normativ bezeichnet dabei die Ausrichtung auf humanistische Ziele wie die Selbstregulation und Autonomie, während sich die neuen Professionalisierungskonzepte auf die Outputsteuerung in Form von Qualitätskontrollen durch Lernstandserhebungen (die Hinweise zur Unterrichtsentwicklung geben sollen und weltweit gängig sind) konzentrieren (Terhart 2012). Grundlegend ist den sich daraus entwickelten Kompetenzmodellen zum einen die Idee einer linearen Wirkungskette, in der eine adäquate Lehrerbildung zu Lehrerkompetenz und diese ihrerseits zu einem Lehrerhandeln im Unterricht führt, das der Unterrichtsqualität zu Gute kommt, die wiederum das Lernen und die Leistungen der Schülerinnen und Schüler fördert (Terhart 2012). Allerdings lässt sich diese Wirkungskette in dieser Form nicht zuverlässig beweisen¹¹, so dass die einzelnen Komponenten lediglich als Einflussfaktoren unter vielen anderen, größtenteils unbekanntem Faktoren aufgefasst werden (Terhart 2012). Zum anderen kommen in einer Weiterentwicklung des Kompetenzmodells persönliche Eigenschaften der Lehrkräfte insofern zum Tragen, als diese auf Lernprozesse innerhalb der Lehrerbildung und damit indirekt auf das kompetente Lehrerhandeln im Unterricht wirken können (Mayr 2014). Dabei stehen sich allerdings wieder zwei grundlegend verschiedene Perspektiven gegenüber: jene, die von einer relativen Stabilität der Persönlichkeit ausgehen, und jene, die von einer relativen Veränderbarkeit der Persönlichkeit ausgehen. Erstere plädieren daher für eine Klärung von Persönlichkeitsmerkmalen vor der Berufswahl zur Lehrerin oder zum Lehrer (mehr dazu in Kap. 2.2.4). Letztere – meist sind hier eher die wissenschaftlichen Ansätze zu verorten – sprechen sich für Möglichkeiten zur Persönlichkeitsentwicklung und deren Formbarkeit im Bildungsverlauf aus und betonen dabei erlernbare Kompetenzen und erlernbare Professionalität (Mayr/Neuweg 2006, Mayr 2014) sowie den Umgang mit beruflichen Belastungen (Terhart 2006). Im Zuge der Erkenntnis der Multidimensionalität und Multikriterialität guten Unterrichts erfolgte seit den 1990ern eine Forschung zum Lehrerberuf, die sich zudem auf Biographien, Zufriedenheit und Beanspruchung im Beruf sowie Geschlechterunterschiede von Lehrerinnen und Lehrern und Lehrerkognitionen konzentrierte, der sich vermehrt zu einem kompetenzorientierten Expertenansatz formte (Haag/Lohrmann 2009).

¹¹ Lediglich im Fach Mathematik gibt es Belege für die Wirkung der Ausbildungsphase auf den Unterricht (Terhart 2012).

Mittlerweile scheint sich in der Unterrichtsforschung verstärkt der Trend abzuzeichnen, sich zur Forschung zum Lehrerberuf und zum Lehrerhandeln zu wandeln, wobei sich der Analysefokus von gesellschaftlichen, institutionellen Rahmenbedingungen (und in diesem Zuge auch zulasten von allgemein-, kultur- und gesellschaftstheoretischen sowie politischen und kritischen Aspekten des Lehrerberufs) hin zu situationsbezogenen, intra- und interpersonellen Faktoren verschoben hat (Terhart 2012):

„Lehrerhandeln im Unterricht: Da Unterrichten weiterhin der Kern der Berufstätigkeit von Lehrkräften ist, bildet das Handeln im Unterricht, seine situativen und intrapersonellen Bedingungen, sein Verlauf im Rahmen der Interaktion mit Schülern sowie seine Wirkungen auf die Schüler, aber auch Rückwirkungen auf Lehrer selbst, ein zentrales Thema der Forschung.“
(Terhart 2012, S. 45)

Diese Feststellung erfolgt parallel zu der weitverbreiteten Annahme von Praktikern, die vornehmlich die Bedeutung der Lehrerpersönlichkeit als den entscheidenden Faktor für Lernerfolge seitens der Schülerinnen und Schüler betonen (Pintarich 2002, Mayr 2014): nicht selten in Form des geborenen Lehrers (vgl. Mayr/Neuweg 2006). Diese Annahme deckt sich mit dem Forschungsbefund, dass sich pädagogisch-didaktisches Erfahrungswissen von Lehrkräften vornehmlich aus dem Repertoire persönlichen Erfahrungswissens zu subjektiven Handlungstheorien formt. Das bedeutet, dass die auf subjektiven Handlungstheorien basierenden Wahrnehmungs- und Handlungsrouninen im Unterricht ebenso wie im *normalen Leben* aus dem persönlichen Erfahrungswissen hervorgehen. Es entstehen dabei eine Art innerer Skripte oder Drehbücher, „denen mehr oder weniger reflektiert gefolgt wird“ (Terhart 2006, S. 45). Folgerichtig fragt daher auch Helmke (2007) unmittelbar: „Was macht den guten Lehrer, was den erfolgreichen Unterricht aus?“ (Helmke 2007, S. 1). Und sicherlich nicht zuletzt konstatierte Hattie: „Auf den Lehrer kommt es an!“ (Hattie nach Terhart 2012, S. 44).

Infolgedessen fühlen sich auch in neueren Werken Autoren dazu berufen, wieder auf die Tradition ideell ausgerichteter Tugendkataloge zurückzugreifen (siehe hierzu bspw. Krenz 2010, Seitz 2008), die sich insgesamt durch tendenziell inhaltsleere Parolen wie „das rechte Maß zwischen Nähe und Distanz“ (Seitz 2008, S. 33) halten oder: „Nur was ich selbst begreife, verstehe und auf mich selbst übertrage, ist ggf. dazu geeignet, als Zielsetzung für Kinder und deren Entwick-

lungsprozess tauglich zu sein.“ (Krenz 2010, S. 199) auszeichnen. Dabei implementieren neuere Tugenden Theorien zu einer fragwürdigen Persönlichkeitserziehung innerhalb des schulischen Rahmens (so bspw. bei Seitz 2008, zur Vertiefung in die Problematik der Persönlichkeitserziehung siehe Paetow 2004 und Sauter 2006) sowie zur Selbstentwicklung der Lehrkraft (vgl. Krenz 2010). Angesichts des völligen Fehlens empirischer Belege (vgl. Bohnsack 2004) sowie der offenen Frage, wie sich das Idealbild des *guten* Lehrers konkret im Handeln ausdrückt (vgl. Terhart 2006), können derartige Plädoyers allerdings weitgehend vernachlässigt werden. Neuerdings haben sich in der Lehrerforschung auch Ansätze entwickelt, die das Vorgehen der traditionellen Lehrerforschung mit dem Versuch des Entwurfs *idealer* Lehrerpersönlichkeiten insofern wieder aufleben lassen, als sie zukünftige Lehrerinnen und Lehrer diesem Raster entsprechend vor- bzw. auszusortieren gedenken (Haag/Lohrmann 2009, mehr dazu in Kap. 2.2.4).

Um sowohl ideellen als auch normativen Ansätzen zur *guten* Lehrkraft entgegen zu wirken, wird Lehrerpersönlichkeit daher hier zunächst¹² mit Mayr und Neuweg definiert als

„Ensemble relativ stabiler Dispositionen, die für das Handeln, den Erfolg und das Befinden im Lehrerberuf bedeutsam sind.“ (Mayr/Neuweg 2006, S. 183)

Im Synergismus der Orientierungen gilt in der Unterrichtsforschung jedoch weiterhin und vornehmlich ein moderner personenzentrierter Ansatz (im Hinblick darauf, dass die Unterrichtsgestaltung bei den Lehrerinnen und Lehrern liegt) unter konstruktivistischem Einbezug der Schülerperspektive auf die Unterrichtsqualität (im Sinne der Klimaforschung) (vgl. Helmke 2006). Hierbei benennt auch Helmke (2009) explizit den Persönlichkeitsbegriff, indem er das Zusammenwirken von Merkmalen der Lehrerpersönlichkeit und der Unterrichtsqualität betont, wobei er allerdings einen entsprechenden Nachweis schuldig bleibt (vgl. Helmke 2009, S. 45f) – wohlgermerkt in allen seinen diesbezüglichen Äußerungen (siehe auch 2004, 2007, 2014). Auch der Ausspruch von Hentigs: „Das wichtigste Curriculum des Lehrers ist seine Person“ (von Hentig 1981, S. 110) wurde ebenso häufig zitiert wie nicht präzisiert. Darüber hinaus wird in Helmkes Angebot-Nutzungs-Modell ersichtlich, dass sich die Wahrnehmung der Schülerin-

¹² In Kapitel 3 wird dieses Verständnis dann anhand des BigFive-Konstrukts weiter ausgebaut und vertiefend erläutert.

nen und Schüler innerhalb der Klassenklimaforschung ausschließlich auf Aspekte des konkreten Unterrichts beziehen und die Schülerinnen und Schüler ihre Lehrerinnen und Lehrer nicht als Menschen oder als Persönlichkeiten wahrzunehmen scheinen. Die direkte Wirkung der Lehrerpersönlichkeit auf die Schülerinnen und Schüler sowie der Zusammenhang dieser Wirkung mit der Wahrnehmung der Unterrichtsqualität bleiben weitgehend unberücksichtigt, da bis dato die Lehrerpersönlichkeit nahezu ausschließlich anhand von Selbstberichten der Lehrkräfte erhoben wurde, die dann mit den Schülerwahrnehmungen verknüpft wurden. Auf diese Weise ließen sich bislang kaum oder nur uneindeutige Eigenschaften *guter* Lehrerinnen und Lehrer ermitteln: Persönlichkeitsmerkmale wie *Extraversion*, *Resilienz* und *Gewissenhaftigkeit* zeigen sich zwar bei sogenannten kompetenten Lehrerinnen und Lehrern, allerdings finden sich diese Merkmale ebenso unter Personen, die in beliebigen anderen Berufen erfolgreich sind (Terhart 2006). Daher schlägt Terhart (2007) in Bezug auf die Lehrerpersönlichkeit vor:

„Man könnte auch Kollegen-, Vorgesetzten- und Elternurteile einholen. Man kann schließlich die Verhaltensweisen derjenigen Lehrer studieren, die von ihren Schülern besonders geschätzt werden und/oder deren Schüler besonders gute Lernfortschritte erzielen.“ (Terhart 2007, S. 20)

Direkte Schülerurteile einzuholen zieht er offenbar nicht in Betracht. Dennoch kommt

„auf der Basis traditioneller wie aktueller Forschung [...] man [...] zu dem Ergebnis, dass es keine spezifischen Persönlichkeitseigenschaften gibt, die eine Person zu einem erfolgreichen Lehrer werden lassen.“ (Terhart 2007, S. 20, vgl. auch Bromme/Haag 2008, Weinert/Helmke 1996)

Hier wird ersichtlich, dass Persönlichkeit im Zusammenhang mit Lehrerinnen und Lehrern als etwas manifest Messbares, objektiv Erfassbares angenommen wird, zu dessen Erfassung Schülerinnen und Schüler offenbar als nicht in der Lage eingeschätzt werden. So konstatiert Terhart (2007) außerdem:

„Sicherlich wirkt sich die Persönlichkeit eines Lehrers auf seine Schüler aus – aber ein und dieselbe Persönlichkeit wirkt auf verschiedene Schüler unterschiedlich, die genaue Wirkungsweise dieses Einflusses ist schwer präzise zu bestimmen.“ (Terhart 2007, S. 20f)

Hier ließe sich allerdings argumentieren, dass es sich mit dem Unterricht ähnlich verhalten mag und sich dennoch ein valides Klassenklima abbilden lässt (vgl. Gruehn 2000). Darüber hinaus zeigen Befunde zum Verhältnis von Fremd- und Selbsteinschätzungen bei Persönlichkeitsmessungen (Kap. 3.4.1), dass Fremdurteile durchaus valide Ergebnisse zeigen können, was Terhart im oben aufgeführten Zitat erwachsenen Beteiligten, wie sogar den Eltern, durchaus zuzutrauen scheint. Ob oder inwiefern die Möglichkeit besteht, die Lehrerpersönlichkeit anhand von Schülerwahrnehmungen zu erheben, soll im Folgenden unter Rückgriff auf die Ausführungen im Kapitel *Perspektivität und Wahrnehmung* (Kap. 1.1) bedacht und im Kapitel *Persönlichkeit* (Kap. 3) dann eingehend beleuchtet werden.

2.1 Operationalisierung des Konstrukts der *guten* Lehrkraft

Ausgangspunkt für die Verknüpfung von Unterrichtsqualität, Lehreffektivität und Lehrperson sind zum einen Forschungen, die nachweisen konnten, dass Schülerevaluationen des Unterrichts aussagekräftig (d.h. reliabel und valide) hinsichtlich bestimmter Aspekte der Lehreffektivität (im Sinne von Lernerfolgen seitens der Schülerinnen und Schüler) sind (bspw. McKeachie 1979); vor allem mehr, als es die Klassengröße oder die individuelle Strenge der Beurteilung verzeichnen können (Cohen 1981). Zum anderen wird basierend auf dem Prozess-Produkt-Paradigma, das davon ausgeht, dass das Lehrerverhalten indirekt über die Unterrichtsgestaltung auf die Lernerfolge der Schülerinnen und Schüler wirkt, untersucht, ob spezifische Prädiktoren speziell im Lehrerverhalten existieren. Dieses Modell beinhaltet zudem Merkmale der Lehrerpersönlichkeit als Vorhersagevariablen, die im Prozess des Unterrichts auf das Verhalten der Lehrkraft einwirken. Das Konstrukt *Persönlichkeit* verweist dabei auf charakteristische, konsistente Muster des Fühlens, Denkens und Handelns (nach Pervin et al. 2005) und gilt als ein relativ valider Vorläufer für das Verhalten (Hogan 1991). Sie kann somit auch als Prädisposition des Verhaltens definiert werden (Guthrie et al. 1998). Daraus schließt bspw. Polk (2006), dass Lehrerpersönlichkeitseigenschaften die Durchführung des Unterrichts beeinflussen, da die Lehrkraft viele Entscheidungen nach ihren Neigungen fällen kann (vgl. auch Rubin/Feezel 1985). Mit anderen Worten: Es wird angenommen, dass das Verhalten respektive das Handeln während des Unterrichtsprozesses in der Persönlichkeit des Lehrers oder der Lehrerin begründet sind.

Bei der Analyse des Zusammenhangs der drei eher abstrakten Basisdimensionen der Unterrichtsqualität (nach Klieme 2006) mit dem konkreten Lehrerhandeln – allerdings ohne Bezug zu diesem Lehrerhandeln möglicherweise vorausgehenden Persönlichkeitsmerkmalen – zeigen sich darüber hinaus hinsichtlich der Lerneffekte seitens der Schülerinnen und Schüler folgende Befunde:

- a) *Klassenführung*: nach Kounin (1976) fließen hier die *Allgegenwärtigkeit, Überlappung, Mobilisierung, Flexibilität* und *Motivierung* der Lehrkraft ein, um Lernzeit zu maximieren und Störungen zu minimieren. Entsprechende Zusammenhänge mit den Lernleistungen der Schülerinnen und Schüler konnten in verschiedenen Studien nachgewiesen werden (Wang/Haertel/Walberg 1993, Rowan/Correnti/Miller 2002, u.a. siehe Lipowsky 1999).
- b) *Kognitive Aktivierung*: die *Herausforderung/Angemessenheit der Aufgaben* wird durch die Lehrkraft bestimmt, um die *fachliche Qualität, die Verarbeitungstiefe* und die *Auseinandersetzung mit dem Gegenstand* zu forcieren. Im Fach Mathematik wurden diesbezüglich schwache bis moderate Effekte auf Lernleistungen nachgewiesen (Shayer/Adami 2007, Klieme 2006).
- c) *Unterrichtsklima*: umfasst die *Lehrer-Schüler- sowie Schüler-Schüler-Beziehungen* und das *Instruktionsverhalten* der Lehrkraft (bzw. die von den beteiligten Schülerinnen und Schülern subjektiv wahrgenommene und geteilte Realität (vgl. Gruehn 2000)). Hier zeigen sich eher indirekte Effekte auf Lernleistungen über die Moderatoren *Motivation* und *Interessenförderung* (Campbell et al. 2004).

Entsprechend hebt Apel (2009) im *Handbuch Unterricht* die Bedeutung der Lehrperson für die Unterrichtsqualität mit folgender Stellungnahme hervor:

„Erfolgreiche Klassenführung beginnt mit dem Auftreten. Vor der Klasse kommt es auf die Körperhaltung, den Blickkontakt, die Körpersprache an.“
(Apel 2009, S. 173)

Im Weiteren konstatiert er:

„Professionelle Klassenführung zeigt sich in kalkulierbarer Zuverlässigkeit und Glaubwürdigkeit, in der Bestimmtheit bei grundsätzlichen Fragen der Disziplin und Großzügigkeit bei kleineren Abweichungen.“ (Apel 2009, S. 173)

Dieser normativen Setzung zufolge verfügten Lehrkräfte erfolgreicher Schulen über ein spezielles *Ethos* im Umgang mit den Schülerinnen und Schülern, anhand dessen sie durch emotionale Stabilität einerseits und wenige, jedoch konsequent

verfolgte Regeln andererseits eine Atmosphäre der Eindeutigkeit und Transparenz schaffen (vgl. Apel 2009).

Helmke (2003) hingegen stellt im Angebot-Nutzungs-Modell die Lehrerpersönlichkeit zwar als Ausgangspunkt der Unterrichtsgestaltung dar. Während dabei die personalen Merkmale außerhalb von Merkmalen der Unterrichtsqualität stehen, die den Unterricht zwar beeinflussen, jedoch kein eigenständiger Aspekt der Unterrichtsqualität sind, ist die Lehrerpersönlichkeit innerhalb dieses Modells nicht direkt als solche durch die Schülerinnen und Schüler wahrnehmbar. Sie tritt lediglich indirekt, vermittelt durch die Unterrichtsqualität zutage. Die Unterrichtsqualität steht innerhalb des Modells für Prinzipien und Merkmale des Unterrichts, die ausschlaggebend für Lernerfolge sind (bspw. Passung, Klarheit, angemessene Methodenvariation, Individualisierung, Motivierung, Klassenführung, effektive Lernzeit, Material – zusammenfassend: Prozessmerkmale des Unterrichts). Bedeutsam ist vor allem Helmkes Feststellung, dass es nicht *den einen (für alle gleich) guten* Unterricht gibt, angesichts der Wechselwirkungen (differenziellen Effekte) zwischen diversen Lehrmethoden und vielfältigen Schülermerkmalen.¹³ Unterschiedlichste *Orchestrierungen* können demnach guter Unterricht sein, d.h. Lernerfolge verzeichnen. Dabei ermöglicht erst der persönliche Ausdruck von Kompetenzen, der weder schematisierbar noch standardisierbar ist, guten Unterricht im Sinne resultierender Lernerfolge seitens der Schülerinnen und Schüler. Auch Weinert (1997) spricht sich gegen pädagogischen Dogmatismus aus, da Bedingungen guten Unterrichts durchaus austauschbar sind. Ebenfalls greifen hier Ansätze, die Lehreffektivität abhängig von der individuellen Passung zwischen Lehrerpersönlichkeit und Unterrichtsmethode interpretieren (vgl. Othmann 2009, Näheres siehe folgendes Kapitel).

Grundlegend für den Kontext dieser Arbeit ist zunächst die Argumentation, dass sich „Unterrichtsmerkmale in Form von Lehrerhandeln“ (Künsting/Billich/Lipowsky 2009, S.657) äußern, ergo Merkmale von Unterrichtsprozessen auf dem Lehrerhandeln gründen. In diesem Sinne stehen Merkmale der Lehrkraft seit jeher im Verdacht sich in Schülerevaluationen zum Unterricht widerzuspiegeln. Um jedoch *gute* Lehrkräfte empirisch abzubilden, bedarf es eines Vorge-

¹³ Eine differentielle Wirkung von Unterricht zeigt sich u.a. am Beispiel des Matthäus-Effekts („Wer hat, dem wird gegeben“): Schülerinnen und Schüler mit günstigen Ausgangsbedingungen profitieren vom Unterricht stärker als benachteiligte.

hens, das über die frühen Methoden der Entwicklung von Tugendkatalogen hinausreicht. So gingen schon Dunkin & Biddle (1974) von sogenannten Presage-Variablen vor allem in Form von Merkmalen der Lehrerpersönlichkeit aus, die Prozess-Variablen des Unterrichts in Form des konkreten Unterrichtsverhaltens der Lehrkraft determinieren, welche ihrerseits wiederum Produkt-Variablen im Sinne von Globalurteilen zur Lehreffektivität beeinflussen. Plausibel erörtern Mayr und Neuweg (2006) die Notwendigkeit zu belegen, dass die Persönlichkeit von Lehrkräften überhaupt Relevanz hinsichtlich ihres Unterrichts besitzt:

„(a) dass es berufsbedeutsame Persönlichkeitsmerkmale gibt (Relevanzthese) und (b) dass diese Merkmale relativ stabil sind (Stabilitätsthese). Andernfalls nämlich wären der Rückgriff auf die das Expertenparadigma kennzeichnenden Konstrukte Wissen oder Können und die Annahme ihrer engen Verkopplung durchaus hinreichend.“ (Mayr/Neuweg 2006, S. 185)

Hier sei angemerkt, dass die Lehrerpersönlichkeit – als Schnittstelle zwischen Persönlichkeitsparadigma und dem Prozess-Produkt-Paradigma der Unterrichtsqualität – in bisherigen Forschungen meist als persönliche Selbsteinschätzung der Lehrkraft entweder mit Schülerleistungen oder mit Beobachter- oder Schülerperspektiven auf in der Regel globale Unterrichtsmerkmale verbunden wurde, was – zumal angesichts Clausens (2002) Befund der perspektivenspezifischen Validität von Wahrnehmung plausiblerweise – zu keinem eindeutigen Ergebnis geführt hat.

Forschungsbefunde zur *guten* Lehrkraft beziehen sich in der Regel auf das im Unterricht zur Schau gestellte Verhalten von Lehrkräften. Diesem wird eine das Lernen der Schülerinnen und Schüler stimulierende Wirkung beigemessen: so beispielsweise dem zum Ausdruck gebrachten Lehr-Enthusiasmus, der vielen Forschungen zur Lehrerpersönlichkeit zugrunde liegt (vgl. Bettencourt et al. 1983, Brigham 1991, Collins 1978, Keller et al. 2013, Kunter et al. 2008), und der sich vornehmlich in Verhaltensweisen wie abwechslungsreicher sprachlicher Darstellung, lebendiger Mimik und Bewegung und somit im Charisma insgesamt äußert. Im engeren Sinne gilt Enthusiasmus allerdings nicht als Persönlichkeitsmerkmal (Rammsayer/Weber 2010, Lang/Lüdtke 2005), kommt jedoch in der Persönlichkeitsforschung in Bezug auf den *Dr.Fox-Effekt* zum Tragen (vgl. hierzu Kap. 2.2.1 und Kap. 3).

Neuere Forschungen operationalisieren die Lehrerpersönlichkeit außerdem in Form der beruflichen Eignung anhand von Prognosen zur Bewährung im Lehramtsstudium. Eine Verwendung entsprechender Ergebnisse findet sich vermehrt innerhalb von Assessment Centern (siehe hierzu das entsprechende Kap. 2.2.4). Solche Assessment Center haben zum Ziel, Studierende vor der Studienwahl zu beraten und mehr oder weniger verbindlich zu selektieren (Asendorpf/Neyer 2012, Hechinger/Seibert 2015). Die BigFive-Persönlichkeitsmerkmale¹⁴ Gewissenhaftigkeit, Extraversion und Neurotizismus zeigen sich in mehreren Studien (Urban 1984, Poropat 2009, Wenzel 2011) als starke Prädiktoren für engagiertes Studierverhalten, Studienleistungen, Kompetenz und Wohlbefinden im Beruf (Mayr/Neuweg 2006, Hanfstingel/Mayr 2007, Mayr 2014). Allerdings ist in Bezug auf diese Befunde zu beachten, dass sie über alle Studienfächer hinweg gelten und nicht explizit für einen Zusammenhang mit der später von den Schülerinnen und Schülern beurteilten Unterrichtsqualität mit den zu Studienbeginn festgestellten Eigenschaften von Lehrkräften sprechen.

2.2 Forschungsüberblick zur *guten* Lehrkraft

Bis zur empirischen Wende innerhalb der Unterrichtsforschung waren Befunde zur Rolle der Lehrkraft hinsichtlich der Lehreffektivität im Allgemeinen unbefriedigend und Nachweise zum Zusammenhang zwischen Lehrermerkmalen und dem Lehrerverhalten fielen uneinheitlich aus. Eher ließen sich zwischen dem Verhalten und den Erwartungen der Lehrkräfte Verbindungen nachweisen (Jansen/Jensen/Mylov 1972). Allerdings konnten Sherman und Blackburn (1975) zeigen, dass Schülerwahrnehmungen der Lehreffektivität stärker mit Lehrerpersönlichkeitsmerkmalen als mit dem Lehrerverhalten verknüpft sind (vgl. auch Hart/Driver 1978), wobei – hier kommt die Schwäche dieser wie vieler anderen Untersuchungen (s.o.) zum Tragen – gänzlich unklar bleibt, aus welcher Perspektive¹⁵ jeweils die Persönlichkeit und das Verhalten erfasst wurden. An den Befund anknüpfend untersuchten Hart und Driver 1978, ob sich Unterrichtsevaluationen je nach

¹⁴ Das Konstrukt der BigFive dient dieser Arbeit als grundlegendes Persönlichkeitsmodell und wird in Kapitel 3.3 präzise vorgestellt.

¹⁵ Generell muss eher davon ausgegangen werden, dass die Persönlichkeit bei nicht explizit genannter Perspektive im Selbstbericht erhoben wird, da innerhalb der empirischen Forschung grundsätzlich davon ausgegangen zu werden scheint, dass dies die sogenannte *wahre* Persönlichkeit darstellt.

Lehrerpersönlichkeit und dies in systematischem Zusammenhang mit der Schülerpersönlichkeit unterscheiden, die beiden letzteren wurden ihrerseits allerdings im Selbstbericht erhoben. Die Persönlichkeitsmessung erfolgte in dieser Studie entlang Jungs (1924) Typologie der Persönlichkeit, die auf der Dichotomisierung von Extraversion und Introversion gründet. Darüber hinaus ordnen sie verschiedene Wahrnehmungsfunktionen (Empfinden vs. Intuition) und Beurteilungsfunktionen (Denken vs. Fühlen) bestimmten Wahrnehmungs- oder Beurteilungstypen zu. Die Unterrichtseffizienz wurde mit 35 Items differenziert erfasst, allerdings bleibt eine Schwachstelle der Untersuchung, dass sie sich auf lediglich 14 Psychologie- oder *Management and Business*-Professoren bezieht und das Geschlechterverhältnis der befragten Studierenden nahezu 1:4 (240 männlich, 928 weiblich) beträgt. Es ließ sich zwar feststellen, dass durchaus generelle Persönlichkeitsmerkmale bei erfolgreichen (Universitäts-)Lehrenden existieren: Extravertierte sowie stärker emotional orientierte Lehrkräfte wurden demnach häufiger positiv bewertet als introvertierte oder rein kognitiv orientierte. Allerdings zeigten sich keine Befunde dahingehend, dass eine Ähnlichkeit zwischen dem Persönlichkeitstypus der Lehrkraft und dem des Lernenden zur Wahrnehmung höherer Lehreffektivität führt oder anderweitig Zusammenhänge zwischen Lehrer- und Schülerwahrnehmungen bestehen.

Erdle, Murray und Rushton (1985) testeten im Weiteren die Hypothese, ob das Lehrerverhalten im Unterricht die Beziehung zwischen Persönlichkeit und Lehrevaluation mediiert. Die Studie basierte auf einer Stichprobe von 37 College-Lehrkräften (davon 35 männlich und zwei weiblich). Die eingesetzten Instrumente erfassten zum einen die Persönlichkeit entlang von 29 Persönlichkeitsmerkmalen¹⁶, beschrieben anhand von Adjektiven mittels derer die Persönlichkeit der Lehrkräfte durch ihre Kollegen eingeschätzt wurde. Zum anderen wurde anhand von Beurteilungen unbeteiligter Beobachter der gezeigte Einsatz von 95

¹⁶ 20 davon wurden definiert durch das *Personality Research Form* (PRF) von Jackson (1974), sieben Merkmale anhand von H.G. Murrays *Study of Personality and College Teaching* (1975) und zudem durch die beiden Merkmale *Neurotizismus* und *Extraversion* nach dem *Eysenck Personality Questionnaire* (EPQ) von Eysenk/Eysenk (1975). Alle Skalen basierten auf einer neunstufigen Likertskalierung.

spezifischen Lehrverhaltensweisen im Unterricht¹⁷ gemessen. Darüber hinaus wurde die Lehreffektivität anhand eines (üblichen) Evaluationsbogens mit einem generellen Globalurteil der Lernenden¹⁸ am Ende des Collegejahres erhoben. Eine Pfadanalyse ergab, dass 50% der Beziehung zwischen Persönlichkeit und Lehreffektivität über das Lehrerverhalten im Unterricht (*Classroom Teaching Behavior*) mediiert werden. Anhand faktorenanalytischer Befunde ließ sich darüber hinaus zeigen, dass Lehrkräfte mit guten (globalen) Schülerevaluationen in zwei Arten persönlicher Orientierungen im Unterricht unterteilbar sind: die *Leistungsorientierung* (v.a. Merkmale, die sich als Dominanz, Intelligenz und Führungsstärke bezeichnen lassen) und die *zwischenmenschliche Orientierung* (Merkmale wie Unterstützung, Nicht-Autoritarismus und Nicht-Defensivität).¹⁹ Außerdem lassen sich zwei Arten an Unterrichtsverhalten identifizieren, nämlich *Charisma* und *Organisation*. Dabei korreliert das Persönlichkeitsmerkmal der *zwischenmenschlichen Orientierung* positiv mit beiden Arten des Unterrichtsverhaltens; das (sogenannte) Persönlichkeitsmerkmal der *Leistungsorientierung* hingegen korreliert jedoch nur mit charismatischem Unterrichtsverhalten. Die Befunde weisen vor allem nach, dass Persönlichkeitsmerkmale und Unterrichtsverhalten durchaus miteinander in Beziehung stehen: Insgesamt weisen 25 von 42 Korrelationen zwischen persönlichen Merkmalen und Unterrichtsverhalten Signifikanz auf. Auch wurde das untersuchte Mediatormodell bestätigt, indem sich 95% der Beziehung zwischen dem Persönlichkeitsmerkmal *zwischenmenschliche Orientierung* und der Unterrichtsevaluation über das Unterrichtsverhalten der Lehrkraft (als Mediator) anhand von *Charisma* und *Organisation* erklären lassen.²⁰ Allerdings beinhaltet das Forschungsdesign ernst zu nehmende

¹⁷ Die Verhaltensweisen wurden nach Murrays *Teacher Behaviors Inventory* (1983) im Unterricht anhand von 3x1 beobachteten Unterrichtsstunden beurteilt – entlang einer fünfstufigen Likertskalierung.

¹⁸ Übersetzt lautet die Frage: Wie würdest du deine Lehrkraft insgesamt in Bezug auf ihre allgemeine Effektivität als Lehrkraft bewerten? Das Antwortformat gab eine fünfstufige Likertskalierung vor.

¹⁹ Die Befunde bestätigen ähnliche Ergebnisse vorhandener früherer Forschungen u.a. von Costin/Grush (1973), Tomasco (1980), Sherman und Blackburn (1975). Die Studien werden in Kap. 2.2.2 detaillierter beschrieben.

²⁰ 34% der Beziehung zwischen dem Persönlichkeitsmerkmal *Leistungsorientierung* und der Unterrichtseffektivität lassen sich über das Unterrichtsverhalten der Lehrkraft (als

kritische Punkte: Erstens ist die Stichprobe mit 37 Lehrenden sehr klein und zweitens nicht repräsentativ, da die evaluierten Lehrenden überwiegend männlich sind und ausschließlich der Fakultät für Psychologie angehören. Drittens wurden die Befunde ausschließlich über Korrelationsmatrizen ermittelt, obwohl hier wesentlich wahrscheinlicher von einer Wirkungsrichtung auszugehen ist – dem Einfluss der Persönlichkeit resp. Persönlichkeitsmerkmalen auf die Unterrichtsbeurteilung²¹. Da der Idee von Persönlichkeit die Idee der situationsübergreifenden und langfristigen Stabilität von Verhalten implizit ist, liegt in Folge der Studie zwar die Vermutung nahe, dass Lehrermerkmale die Lehrevaluation beeinflussen (Erdle/Murray/Rushton 1985); die Ergebnisse jedoch sind – wie es durchweg für derartige Untersuchungen zum Thema Lehrerpersönlichkeit gilt – aufgrund ihrer korrelativen Anlage nicht gut interpretierbar und nicht generalisierbar. Dennoch lässt sich Lehrerverhalten im Unterricht, das positiv mit Globalurteilen zur Lehre korreliert, in zwei Bereiche aufspalten: *Charismatisches Verhalten* (Enthusiasmus, Harmonie, Interessantheit von Inhalten, Partizipation) und *organisatorisches Verhalten* (gute Vorbereitung, Strukturiertheit, Klarheit). Dieser Befund weist Ähnlichkeiten zu Freys (1978) zweidimensionalem Modell anhand von Faktorenanalysen über verhaltensorientierte Studentenevaluationsinstrumente auf (vgl. Erdle/Murray/Rushton 1985) und verweist vor allem auf die differenzierte Wahrnehmung der Schülerinnen und Schüler zwischen unterrichtlichem und persönlichem Verhalten im Unterricht.

In seiner Metaanalyse fasste Feldmann (1986) die bis dato untersuchten Korrelationen zwischen der persönlichen Art des Unterrichtens seitens der Lehrkraft und entsprechenden Lerneffekten seitens der Schülerinnen und Schüler zusammen. Dazu generierte er eine Sammlung von 14 persönlichen Merkmalen, denen er globale Effektivitätseinschätzungen des Unterrichts (*Teacher Effectiveness in the Classroom*, *Instructional Effectiveness* und *Teacher Effectiveness*) gegenüberstellte. Dabei konnte er feststellen, dass Persönlichkeitsselbsteinschätzungen der Lehrerinnen und Lehrer mit den Schülereinschätzungen der Unterrichtseffektivität eher wenig korrelieren: Vier der erfassten Persönlichkeitsmerkmale

Mediator), das *Charisma*, erklären. Insgesamt können 50% der geteilten Varianz von Persönlichkeit und Effektivität durch charismatisches und organisiertes Verhalten im Unterricht erklärt werden.

²¹ Was jedoch auch andersherum denkbar ist: dass erfolgreicher Unterricht persönlichen Lehrereigenschaften zugeschrieben wird, die dann wiederum im vermehrten Einsatz zu erfolgreichem Unterrichten führen (vgl. Erdle/Murray/Rushton 1985).

korrelierten dabei schwach mit der Unterrichtseffektivität, zwei davon sogar sehr schwach (mit 0,27 und 0,30). Beobachtereinschätzungen (sowohl aus Schüler-/Schülerinnen- als auch Kollegenperspektive) der Lehrerpersönlichkeit korrelieren mit den Effektivitätseinschätzungen der Schülerinnen und Schüler dagegen bei elf der 14 Merkmale moderat bis stark (0,50 und höher). Ein Fazit Feldmanns ist daher, dass es hinsichtlich der Bedeutung des Zusammenhangs von wahrgenommenen Merkmalen der Lehrerpersönlichkeit angebracht scheint, verschiedene Perspektiven auf die Lehrerpersönlichkeit getrennt voneinander zu analysieren. Dabei attestiert er den Wahrnehmungen der Schülerinnen und Schüler zwar Validität, betont jedoch, dass es sich dabei nicht um die Beschreibung einer *wahren* Persönlichkeit handelt, sondern eben lediglich um eine spezifische, zumal kontextbezogene Perspektive, die allerdings angesichts der Möglichkeit der Aggregation auf Klassenebene (die auch der Lehrerebene entspricht) eine höhere Reliabilität aufweist und weniger verzerrt ist. Darüber hinaus zeigten seine Analysen, dass Schülerinnen und Schüler zwischen Unterrichtseffektivität und Lehrerpersönlichkeit differenzieren. Allerdings könnten Persönlichkeitseinschätzungen sowie Effektivitätseinschätzungen durch die Schülerinnen und Schüler denselben Verzerrungsvariablen und/oder Erwartungen unterliegen und damit die Höhe der Korrelationen bedingen (Feldmann 1986).

Über den Befund von Marsh und Hocevar (1991) hinaus, dass parallel zur zeitlichen Stabilität von Persönlichkeitsmerkmalen auch Schülerurteile des Unterrichts bestimmter Lehrerinnen und Lehrer über lange Zeiträume (13 Jahre) relativ stabil sind, ließen sich kaum Zusammenhänge von Unterrichtsevaluationen durch Schülerinnen und Schüler mit Lehrerselbsteinschätzungen der Persönlichkeit verzeichnen (Bendig 1953, Sorey 1968). Wohl aber ließen sich Zusammenhänge von Unterrichtsevaluationen durch Schülerinnen und Schüler mit Fremdeinschätzungen durch Lehrerkollegen in Bezug auf verschiedene Eigenschaften nachweisen: So zeigten sich positive Zusammenhänge von Schülerevaluationen mit fremdeingeschätzten Eigenschaften wie Führung, Objektivität, wenig Angst, Extraversion, Aufgabenorientierung, Unterstützung, Nicht-Autoritarismus (Murray 1975, 1978, 1983). Resümierend muss hierzu jedoch festgehalten werden, dass sowohl die Methoden als auch die statistischen Verfahren der empirischen Schulforschung aus dieser Zeit in weiten Teilen unzulänglich sind und schwer interpretierbare Resultate hinterlassen, zumal sie vornehmlich korrelativer Art sind. Erschwerend kommt außerdem hinzu, dass in der Verknüpfung dieser beiden in sich schon uneinheitlichen Paradigmen, der Persönlichkeitsforschung und der

Prozess-Produkt-Forschung, ein offensichtliches Chaos herrscht und sich Methoden, Perspektiven und Ansätze in unsystematischer Weise vermischen. In Folge dieser Befunde verstärkte sich allmählich die Differenzierung zwischen den Perspektiven bei Erhebungen von Persönlichkeitsmerkmalen. Murray, Rushton und Paunonen (1990) ließen demnach in ihrer Untersuchung von 46 Universitätsprofessoren der Fakultät Psychologie deren Persönlichkeit anhand von 29 Persönlichkeitsmerkmalen, u.a. nach Eysencks *Neurotizismus* und *Extraversion* durch Kollegen einschätzen. Diese Fremdeinschätzungen der Lehrerpersönlichkeit wurden mit allgemeinen, globalen Effektivitätsbeurteilungen von ihren Studierenden aus sechs verschiedenen Kursarten (mit unterschiedlichen Teilnehmenden, Teilnehmerzahlen und Methoden) in Zusammenhang gebracht. Dabei zeigt sich, dass die Effektivitätsurteile derselben Professoren je nach Kursart variierten, wobei kaum einer überall gut oder überall schlecht abschneidet. Darüber hinaus konnten Unterrichtsevaluationen durch Studierende anhand der Kollegeneinschätzung der Persönlichkeit für jede Kursart vorausgesagt werden, wobei sich insgesamt die Merkmale Extraversion, Führung, Liberalismus, Unterstützung, Intelligenz und Wandelbarkeit als bedeutsam erwiesen. Die spezifischen Persönlichkeitsmerkmale, die zur empfundenen Lehreffektivität beitragen, differieren zudem deutlich je nach Kursart. D.h. Kurse mit jüngeren Studierenden assoziieren andere Eigenschaften mit Effektivität als höhere Kurse. Die Ergebnisse werden von den Autoren dahingehend interpretiert, dass nicht jede Lehrerpersönlichkeit für jede Art Unterricht geeignet ist, es also einer Passung von Persönlichkeit und Methode bedarf. Eine alternative Interpretation ist zudem, dass die Persönlichkeit sich in jenem Verhalten äußert, das den jeweiligen Unterricht prägt (siehe hierzu Erdle/Murray/Rushton 1985). Das Plädoyer für eine Vorauswahl von Lehrkräften entweder durch Persönlichkeitsselektion oder durch Zuordnung zu entsprechenden Kursarten findet mit dieser Studie eine erste empirische Grundlage. Viele Studien bestätigten in den nächsten Jahren die mittels Peer-Ratings²² erfasste Lehrerpersönlichkeit als signifikanten Prädiktor für effektiven Unterricht bzw. für positive Schülerratings bzgl. der Lehreffektivität (Schmidt et al. 1991, Murray et al. 1990, Magno/Sembrano 2008). So zeigten bspw. Renaud und Murray (1996) in einer Untersuchung, die auf Peer-Ratings

²² Überwiegend ist hier gegenseitiges Peer-Rating gemeint: dieses gilt als intern konsistenter und stärker mit Lehreffektivität korreliert als Selbst- oder Schülereinschätzung der Lehrerpersönlichkeit (Feldmann 1986).

von 33 Psychologieprofessoren einer Fakultät beruht, dass die Lehreffizienz signifikant mit einzelnen Persönlichkeitseigenschaften bzw. -attributionen (Führung zeigend, Zustimmung suchend, ordentlich, Aufmerksamkeit suchend, unabhängig, gesellig) korreliert, nicht aber mit den BigFive-Dimensionen *Extraversion* oder *Neurotizismus* (Näheres zu den BigFive-Dimensionen siehe Kap. 3.3).

„The personality traits of the teacher can greatly influence student ratings of instructor effectiveness. [...] personality traits taken collectively can account for up to 75% of between-teacher variance in student instructional ratings.”

(Murray 1996, S. 324, vgl. Feldmann 1986, Murray et al. 1990)

Allerdings weist auch diese Untersuchung Schwächen im Design auf: so wurden zehn Unterrichtsmerkmale anhand von zehn Einzelitems²³ erfasst, woraus ein Generalfaktor *Lehreffektivität* erstellt wurde. Die Persönlichkeit hingegen wurde anhand von 29 Persönlichkeitsmerkmalen in einer Mischung aus verschiedenen Inventaren erfasst (bspw. Jackson 1974, Murray 1983, Eysenck/Eysenck 1975); sie lässt demnach als uneinheitlich bezeichnen.

Insgesamt steht den persönlichkeitsfokussierenden Ansätzen ein Ansatz gegenüber, demzufolge effektiver Unterricht davon abhängt, wie eine Lehrkraft die Methoden für ihren Unterricht auswählt. Da aus Schülersicht die Unterrichtsqualität von ein und derselben Lehrkraft fachübergreifend variieren kann – wobei gewisse persönliche Merkmale sich in einem Fach als hilfreich, im anderen als belastend auswirken können (Murray et al. 1990) – kommen Norzila, Parilah und Fauziah (2007) zu dem Schluss, dass es auf die Angemessenheit der Methode im Sinne einer Passung zwischen Unterrichtsstil und Persönlichkeit ankommt, um effektiv zu lehren. Die meisten Studien zum Zusammenhang von Lehreffektivität und Lehrerpersönlichkeit stammen aus dem angloamerikanischen Raum. Eine ihrer wesentlichen Schwächen besteht fast durchweg darin, dass die Perspektivität der Einschätzungen nicht differenziert reflektiert wird. Oft scheint es nicht einmal erwähnenswert, wessen Einschätzungen resp. Perspektiven den Ergebnissen zugrunde liegen. Dies deutet darauf hin, dass hier innerhalb des Persönlichkeitsparadigmas noch immer von einer objektiven, wahren Persönlichkeit ausgegangen wird, was keine Basis für eine systematische Erhebung darstellt.

²³ Die Erfassung erfolgte über eine 5-stufige Likertskalierung.

Eine weitere Schwäche der meisten Studien zu dem Thema – die außerdem häufig innerhalb von Universitäten und selten in Schulen erfolgen – ist, dass die Messung der Effektivität nicht etwa in Form von längsschnittlichen Lernleistungstests, sondern in Effektivitätseinschätzungen der Studierenden als Äquivalent zu globalen, allgemeinen Einschätzungen zur Unterrichtsqualität erfolgt. Es besteht dabei jedoch die Möglichkeit, dass Studierende resp. Schülerinnen und Schüler durchaus von Lehrkräften lernen (also nachweislich über einen Lernzuwachs verfügen), die sie selbst nicht als besonders effektiv einschätzen. Entsprechend stellte Helmke (1996) bzgl. studentischer Lehrevaluation heraus, dass die subjektive Zufriedenheit zwar mit dem selbstempfundenen Lernerfolg korreliert, aber nicht mit nachweislichen Lernerfolgen. Daher bedürfte es ähnlich aufgebauter Studien, jedoch mit einer anderen Evaluationsform der Effektivität sowie der gezielten Einbindung der Perspektive auf die Lehrerpersönlichkeit. Entsprechend zeigen Seidel und Shavelson in ihrer Metaanalyse von *Teaching-Effectiveness-Studies* (2007) zwar Zusammenhänge von Lehrerhandeln und Schulerfolg auf, kritisieren jedoch grundlegend die Inkonsistenz und mangelnde Güte vieler Studien und plädieren für ein Research-Based-Design.

Trotz einer uneindeutigen Befundlage konstatiert Lipowsky (2007), dass empirische Befunde vermehrt auf die Bedeutsamkeit der Lehrperson für die Lernentwicklung der Schülerinnen und Schüler hinweisen, v.a. die Value-Added-Studien aus den USA: auf Klassenebene aggregiert ergeben deren Residuen Indikatoren für die Teacher Quality (oder den Classroom Effectiveness Index), die stärker zu wirken scheint als bislang angenommen. Babu und Mentro präsentierten 2003 auf der AERA-Tagung entsprechende Ergebnisse, denen zufolge Rechen- und Lese-Leistungen jüngerer Schülerinnen und Schüler stärker unter jenem Classroom Effectiveness Index der Lehrkraft leiden als ältere, außerdem leistungsstarke Schülerinnen und Schüler (mit guten Lernvoraussetzungen) weniger als leistungsschwächere mit ungünstigen Lernvoraussetzungen. Hinzu kommt, dass Klassenmerkmale einen stärkeren Einfluss auf die Lernleistung aufweisen als Schulmerkmale (Hosenfeld/Helmke/Ridder/Schrader 2001, Hattie 2013, Value-Added-Studien (Längsschnitt): Babu/Mentro 2003, Lanahan et al. 2005). Entsprechend stellt Hattie (2013) in seiner Metaanalyse mehrerer Value-Added-Studien heraus, dass 30% der Leistungsunterschiede zwischen den Schülerinnen und Schülern auf Lehrermerkmale – wie Wissen und Handeln – zurückzuführen sind, welche damit den größten Einfluss darstellen (nach Schule, Peers, soziale Herkunft) (siehe auch Rowan/Correnti/Miller 2002; Lanahan et al. 2005). Value-

Added-Studien geben allerdings keine differenzierten Hinweise auf bestimmte, relevante Merkmale der Teacher Quality, da sie anhand allgemeiner, globaler Einschätzungen erhoben werden.

2.2.1 Forschungsbefunde zur Lehrerpersönlichkeit

Nach der Feststellung, dass die Lehrerpersönlichkeit insgesamt überraschend wenig systematisch erforscht ist²⁴, nahm Wachtel (1998) erneut eine umfassende Untersuchung von Lehrermerkmalen in Angriff. Dabei konnten keine klaren Befunde zu persönlichen Merkmalen wie Alter und Erfahrung der Lehrkräfte herausgestellt werden. Im Durchschnitt korrelieren jedoch schlechte Reputationen und ein schlechter Ruf der Lehrkräfte mit schlechten Beurteilungen ihres Unterrichts, wobei wiederum die Selbsteinschätzung der Lehrerpersönlichkeit wenig mit Schülerfeedbacks zum Unterricht, die Schülereinschätzungen von Lehrermerkmalen allerdings signifikant mit den Unterrichtsfeedbacks zusammenhängen. Im Weiteren wurde jedoch nicht abschließend unterschieden, ob es sich bei den Zusammenhängen um einen validen Einfluss oder um Biasfaktoren handelt (vgl. auch Feldmann 1986). Zudem weist das Geschlecht meist keinerlei signifikante Unterschiede auf, allerdings wird das eigene Geschlecht etwas häufiger positiv bewertet (vgl. auch Feldmann 1993).

In neuen Studien entlang eines um das Persönlichkeits-Paradigma ergänzten Angebot-Nutzungs-Modells, das von bestimmten Eigenschaften erfolgreicher Lehrerinnen und Lehrer²⁵ ausgeht, zeigten sich entgegen der Befunde aus Feldmanns Metaanalyse (1986) nur schwache Korrelationen zwischen der Lehrerpersönlichkeit und direkten Effekten in Form von Schülerleistungen. Daher gehen deren Autoren weiterhin von einem starken Mediatoreffekt aus (siehe hierzu Helmke 2003, Lipowsky/Rakoczy/Klieme/Reusser/ Pauli 2005). Aktuellere Studien fo-

²⁴ Diese Feststellung wiederholt sich im Laufe der Zeit: Sie wurde schon 1981 nach eingehender Literaturprüfung von Seiffke-Krenke (siehe Saldern 1987) getroffen und erneut 2006 von Clayson und Sheffet (vgl. Kap. 2.2.2).

²⁵ Eigenschaften stehen dabei im Gegensatz zum Experten- oder Novizenparadigma, das auf Wissen und Überzeugungen, sprich: kognitiven Aspekten der Lehrkraft (Lehrerkognitionen) basiert und denen sowohl theoretisch als auch empirisch eine Bedeutung für die individuelle Unterrichtsgestaltung zugesprochen wird.

kussieren zudem konkretere Eigenschaften wie das Lehrerselbstwirksamkeitserleben, den Optimismus, die Belastungsfähigkeit²⁶ und den Enthusiasmus, die mit dem Berufsalltag sowie der Lehrer-Schüler-Beziehungen und darüber dann mit den Schülerleistungen in Zusammenhang gebracht werden (siehe u.a. Dworkin 1997, Stoeckli 2000, Schaarschmidt 2001, Klusmann et al. 2006, Kunter et al. 2008, Baier et al. 2018). Die Ergebnisse sind zum einen äußerst uneinheitlich. Zum anderen bilden die Konstrukte der Studien keine Persönlichkeitseigenschaften per definitionem ab, sondern erfassen eher das individuelle und subjektive Empfinden und Erleben der Lehrpersonen, was nicht identisch mit dem sichtbaren und von den Schülerinnen und Schülern beobachtbaren Verhalten sein muss (dieser Aspekt wurde in Kap. 1.1 ausführlich diskutiert).

In einer Befragung von Oberstufenlehrern und -lehrerinnen zu deren Lehrverhalten und persönlichen Eigenschaften²⁷ (Randoll 1995) verneint die Mehrzahl der Lehrkräfte Unsicherheiten gegenüber den Schülerinnen und Schülern, was sich

²⁶ Klusmann et al. (2006) etwa bildeten im Fach Mathematik einen Zusammenhang zwischen beruflichem Belastungserleben und durch die Schülerinnen und Schüler wahrgenommener Unterrichtsqualität ab. Dabei erhoben sie das Belastungserleben der Lehrerinnen und Lehrer auf der Basis von AVEM: Die vier AVEM-Typen resp. Muster basieren auf der Differenzierung zwischen Gesundheitstyp, Schonungstyp und zwei unterschiedlichen Varianten an Risikotypen. Die Unterrichtsqualität aus Schülersicht wurde anhand von sechs Prozessmerkmalen erhoben: Störungen im Unterricht, Interaktionstempo, kognitive Selbstständigkeit, Sozialorientierung der Lehrkraft, Gerechtigkeitsorientierung, Kränkung durch die Lehrkraft) ab; als Synonym zur Unterrichtsqualität wird dabei global der Begriff des Lehrerverhaltens im Unterricht verwendet. Ausgangspunkt der Untersuchung ist die Frage nach möglichen, durch die Schülerinnen und Schüler beobachtbaren Auswirkungen des arbeitsbezogenen Erlebens und Verhaltens von Lehrkräften in ihrem Unterricht. Auch in jener Untersuchung werden die personalen Einschätzungen (wie es der AVEM jedoch eher erfordert als die BigFive) durch die Lehrkräfte selbst erfasst und dann perspektivenübergreifend in Zusammenhang mit der schülerperzipierten Unterrichtsqualität sowie den Schülerleistungen gebracht. Zum einen zeigen sich bei einer ein-faktoriellen Varianzanalyse zwischen den AVEM-Typen signifikante Unterschiede in vier der sechs Prozessmerkmale: in Bezug auf das Interaktionstempo, die kognitive Selbstständigkeit, die Sozialorientierung und die Gerechtigkeitsorientierung wird der Gesundheitstyp signifikant positiver als die anderen Typen wahrgenommen. Zum anderen lassen sich auf Klassenebene keine direkten regressionsanalytischen Effekte der Lehrertypen auf die Schülerleistungen ausmachen. Das Belastungserleben von Lehrkräften, das als Grundlage für die gesundheitliche Bewältigung beruflicher Beanspruchungen gilt (siehe hierzu auch Kap. 2.2), lässt demnach keinen eindeutigen Rückschluss auf die Lernleistungen der Schülerinnen und Schüler zu.

²⁷ Der Fokus der Studie lag vornehmlich auf Unterschieden der Wahrnehmung zwischen ost- und westdeutschen Lehrkräften.

einerseits als eine gewisse emotionale Stabilität interpretieren ließe. Andererseits schätzen sich etwa 28% als ungeduldig gegenüber den Schülerinnen und Schülern ein, was gemäß den BigFive als Merkmal der Dimension *Neurotizismus* resp. *Resilienz* gilt (siehe hierzu vertiefend Kap. 3.3). Insgesamt fallen die Selbsteinschätzung bzgl. Kompetenz, Fähigkeiten und Persönlichkeit resp. Verhalten sehr positiv aus, bspw. schätzt sich mit ca. 80% der überwiegende Teil der befragten Lehrkräfte als selbstkritisch und in der Lage, Fehler zuzugeben, ein. Ein für diese Arbeit herausragender Befund der Lehrerbefragung von Randoll (1995) ist allerdings, dass 31-40%²⁸ der Lehrerinnen und Lehrer aussagen, dass sie den Schülerinnen und Schülern gegenüber nicht offen zeigen, „was sie wirklich denken und fühlen“ (Randoll 1995, S. 21). Der überwiegende Teil jedoch zeigt demgegenüber zumindest meistens „was sie denken und fühlen“ (Randoll 1995, S. 21). Zudem wurde in weiteren Untersuchungen nachgewiesen, dass das Alter von Lehrkräften signifikant negativ mit deren *Neurotizismus* und positiv mit ihrer *Extraversion* korreliert (Murray 1996), wobei dies auf generelle Schwankungen persönlicher Eigenschaften mit dem Alter zurückzuführen ist (vgl. Kap. 3.4.3). Auch Mayr und Neuweg (2006) befassen sich erneut mit der Stabilität der Lehrerpersönlichkeit, um damit verstärkt für einen neu ausgerichteten Fokus auf die Lehrerpersönlichkeit über die reine Professionalisierung hinaus zu argumentieren (*Stabilitätsthese*). Es konnten dabei Veränderungen der Merkmale innerhalb eines relativ stabilen Rahmens gezeigt werden.²⁹

Hinsichtlich ihrer Befunde zur Lehrerpersönlichkeit und deren Interpretierbarkeit ist die Untersuchung von Mayr und Neuweg (2006) insgesamt am einschlägigsten: 60 Lehrerinnen und Lehrer einer betriebswirtschaftlichen Handelsakademie wurden explizit aus Lehrer- und Schülersicht in ihrer Persönlichkeit entlang der BigFive-Dimensionen (vgl. hierzu Kap. 3) sowie im unterrichtlichen Verhalten der Lehrkräfte entlang schülerorientiert-kommunikativem versus kontrollierendem Verhalten eingeschätzt. Die Ergebnisse belegen deutliche perspek-

²⁸ Gemäß dem Studiendesign werden Ost- und Westergebnisse differenziert berichtet. Demnach gilt der folgende Befund für 31% der befragten Lehrerinnen und Lehrer aus Ostdeutschland und 40% der befragten Lehrerinnen und Lehrer aus Westdeutschland (Randoll 1995).

²⁹ Allerdings wurden diesbezüglich wieder nur Selbsteinschätzungen der Lehrkräfte (entlang der Merkmale Kontaktbereitschaft, Stabilität und Selbstkontrolle) untersucht.

tivenspezifische Korrelationen, d.h. Persönlichkeits- und Unterrichtseinschätzungen aus derselben Perspektive weisen deutlich mehr und – zumindest in der Schülerperspektive – auch eher höhere Korrelationen auf als das *klassische* Untersuchungsmodell, das Persönlichkeitsselbsteinschätzungen der Lehrkräfte mit Unterrichtseinschätzungen durch die Schülerinnen und Schüler kombiniert.

<i>Persönlichkeit aus Lehrersicht</i>	<i>Schülerorientiert-kommunikatives Verhalten aus Lehrersicht</i>	<i>Kontrollierendes Verhalten aus Lehrersicht</i>
Neurotizismus	-,28*	
Extraversion	,53**	,30*
Offenheit	,41**	,28*
Verträglichkeit		
Gewissenhaftigkeit	,46**	,22

<i>Persönlichkeit aus Lehrersicht</i>	<i>Schülerorientiert-kommunikatives Verhalten aus Schülersicht</i>	<i>Kontrollierendes Verhalten aus Schülersicht</i>
Neurotizismus		
Extraversion	,45**	
Offenheit	,41**	
Verträglichkeit		
Gewissenhaftigkeit		

<i>Persönlichkeit aus Schülersicht</i>	<i>Schülerorientiert-kommunikatives Verhalten aus Schülersicht</i>	<i>Kontrollierendes Verhalten aus Schülersicht</i>
Neurotizismus	-,66**	-,46**
Extraversion	,84**	
Offenheit	,76**	
Verträglichkeit	,62**	
Gewissenhaftigkeit	,30*	,64**

60 Lehrer/innen des Faches Betriebswirtschaft an Handelsakademien und deren Schüler/innen (Schülerdaten aggregiert auf Klassenebene); Pearson-Korrelationen, die zumindest eine Tendenz zur Signifikanz aufweisen ($p < .10$), * $p < .05$, ** $p < .01$ (zweiseitig)

Abb. 1: Zusammenhänge zwischen Persönlichkeit und Verhalten von Lehrkräften (aus Mayr/Neuweg 2006, S. 191).

Es zeigt sich zudem, dass der Zusammenhang von persönlichen Lehrermerkmalen mit unterrichtlichem Lehrerverhalten je nach Charakter bzw. Art der Verhaltensebene variieren. Dies zieht den Schluss nach sich, dass Schülerinnen und Schüler durchaus in ihren Einschätzungen differenzieren. Interessant ist, dass die

Gewissenhaftigkeit, die im Rahmen der Lehrerausbildung als wichtige persönliche Eigenschaft gilt (siehe hierzu Kap. 2.2.4), aus Schülerperspektive nicht in jeder Hinsicht von herausragender Bedeutung ist (vgl. Mayr 2003) – sie ist in Bezug auf schülerorientiert-kommunikatives Verhalten eher schwach ausgeprägt. Mayr und Neuweg schließen angesichts der Befunde „denkbare unmittelbare Wirkungen der Lehrer/innen/persönlichkeit auf die Schüler/innen nicht aus“ (Mayr/Neuweg 2006, S. 192). Mit diesem Befund verstehen sie die *Relevanzthese* (s.o.) als belegt.

Die neueste Studie von Mayr und Neuweg (2018), fokussiert hingegen auf eine konstruktivistische (entgegen einer instruktivistischen) unterrichtsmethodische Grundeinstellung der Lehrkräfte in Zusammenhang mit Persönlichkeitsmerkmalen. Darin wird zum einen nachgewiesen, dass eine konstruktivistische Einstellung der Lehrkraft nur bedingt mit gesteigerten Lernleistungen, jedoch aber mit einer durch die Schülerinnen und Schüler positiveren Wahrnehmung des Unterrichts einhergeht. Zum anderen lassen sich moderate signifikante Zusammenhänge zwischen einer ausgeprägt konstruktivistischen Grundeinstellung und den selbsteingeschätzten BigFive-Persönlichkeitsdimensionen *Extraversion* und *Offenheit für neue Erfahrungen* (explizite Ausführung zu den BigFive-Dimensionen in Kap. 3.3) sowie etwas abgeschwächer mit dem Lebensalter und der Dauer der Lehrpraxis nachweisen.

Baier et al. (2018) weisen mit ihrer Untersuchung³⁰ hingegen nach, dass eine konstruktivistische Grundeinstellung der Lehrkraft keine Effekte auf die durch die Schülerinnen und Schüler eingeschätzte Unterrichtsqualität entlang der drei Basisdimensionen nach Klieme (2006, vgl. Kap. 2.1) zu haben scheint. Die Bedeutung einer konstruktivistischen Grundeinstellung für die Unterrichtsqualität bleibt demnach fraglich. Allerdings wird in der Studie gezeigt, dass von den durch die Lehrkräfte selbsteingeschätzten BigFive-Dimensionen *Extraversion* und *Gewissenhaftigkeit* zwar lediglich schwache, jedoch moderat signifikante Effekte (0,22-0,26) auf die Schülereinschätzungen der drei Basisdimensionen

³⁰ In der Studie wurden diverse lehrpersonenbezogene Merkmale aus Forschungsbefunden extrahiert, die sich als unterrichtsqualitätsrelevant erwiesen haben. Diese einzelnen Merkmale (kognitive Fähigkeiten, Extraversion, Gewissenhaftigkeit, professionelle Kompetenz, konstruktivistische Grundeinstellung, Enthusiasmus) wurden wiederum innerhalb eines multivariaten Regressionsmodells im Hinblick auf ihre jeweiligen Effekte auf die schülerperzipierte Unterrichtsqualität (entlang der drei o.g. Basisdimensionen) geprüft. Insgesamt folgt die Studie einem eher explorativen Ansatz.

der Unterrichtsqualität nach Klieme (2006) haben. *Extraversion* wirkt dabei positiv auf die Unterrichtsqualität, während die Effekte der *Gewissenhaftigkeit* negativ ausgeprägt sind. Die Autoren kommen demnach zu dem Schluss, dass die selbsteingeschätzte *Extraversion* und *Gewissenhaftigkeit* der Lehrkräfte auf deren durch die Schülerinnen und Schüler eingeschätzte Unterrichtsqualität wirken (Baier et al. 2018).

Auch Othmann (2009) untersucht dahingehend ähnliche Modelle wie die vorliegende Arbeit, als sie die Wirkung von BigFive-Persönlichkeitsdimensionen³¹ auf den Unterricht analysiert. Die Studie basiert auf Daten von 391 Lehrerinnen und Lehrern aus verschiedenen Regionen Malaysias. Allerdings erfolgt in ihrer Studie die Persönlichkeitsmessung ausschließlich anhand von Lehrerselbsteinschätzungen der BigFive. Zudem ist das aus Schülersicht erhobene Konstrukt der Lehreffektivität – in Anlehnung an McKeachie (1979) als Grad des Erreichens anvisierter Bildungsziele verstanden – keines, das die Unterrichtsqualität ebenso adäquat und differenziert widerzuspiegeln vermag wie die verschiedenen Prozessmerkmale des Unterrichts. Sie weist in ihrer Untersuchung signifikante Korrelationen der Persönlichkeitsdimensionen *Extraversion*, *Verträglichkeit* und *Gewissenhaftigkeit* mit der Lehreffektivität nach, jedoch keine Beziehungen zwischen Lehreffektivität und *Neurotizismus* und *Offenheit*. Trotz der hohen Korrelationen zeigen sich in Regressionsanalysen allerdings kaum Wirkungseffekte von Persönlichkeitseigenschaften auf die Lehreffektivität. Othmann resümiert abschließend, dass Lehreffektivität auf zweierlei Kriterien beruht: erstens auf den eingesetzten Unterrichtsmethoden und zweitens auf der Passung zwischen Unterrichtsmethoden und Lehrerpersönlichkeit. Auch Kneipp, Kelly, Biscoe und Richard (2010) untersuchen den Zusammenhang von Persönlichkeitsselbsteinschätzungen (entlang der BigFive) von 63 Lehrenden an der Universität mit Studentenevaluationen zur Unterrichtseffektivität (aus dem Archiv der letzten zwei Jahre). Im Ergebnis zeigt sich, dass *Verträglichkeit* die einzige Dimension ist, die signifikant mit den Studentenurteilen korreliert und eine stärkere Prädiktionskraft als die anderen Merkmale aufweist.

Hinsichtlich des Zusammenhangs von Lehrerpersönlichkeit und Unterrichtsqualität offenbart sich insgesamt eine ausgesprochen uneinheitliche Forschungslage.

³¹ Vertiefend zu den BigFive-Persönlichkeitsdimensionen siehe Kap. 3.3.

Nachdem die Persönlichkeitsforschung innerhalb der Schul- und Unterrichtsforschung lange Zeit in Verruf geriet, zeigt sich jedoch in entsprechenden Forschungen jüngerer Zeit, dass auf die Einbindung der Lehrkraft als Person in Modellen zur Unterrichtsqualität nicht gänzlich verzichtet werden kann, was zu diversen Differenzierungen und einer unübersichtlichen Handhabung von Begrifflichkeiten – wie bspw. Lehreffektivität, Lehrkompetenz, Lehrerpersönlichkeit, Lehrermerkmale, Lehrerhandeln, Lehrerverhalten etc. – führte.

2.2.2 Dr.Fox-Effekt

Schon früh (seit etwa den 1970ern) standen die Forschung zur Lehrerpersönlichkeit sowie die Forschung zur Unterrichtsqualität aus Schülersicht unter dem Verdacht, durch den Dr.Fox-Effekt erheblich verzerrte Ergebnisse abzubilden. Im Experiment, auf dem der Dr.Fox-Effekt basiert, unterrichtete ein Schauspieler äußerst charismatisch und substanzlos zu einem Thema, über das er nichts wirklich wusste. Selbst erfahrene Ausbilder fühlten sich danach hinsichtlich ihres Lernerfolgs zufrieden gestellt. Dieser gut untersuchte Effekt besagt, dass ein blenderisches, enthusiastisches Auftreten und charismatisches Verhalten einer Lehrperson dazu führt, dass die Zuhörerinnen und Zuhörer bzw. Schülerinnen und Schüler dieses Auftreten mit Kompetenz und dem Gefühl, etwas Wesentliches zu lernen, verknüpfen – unabhängig vom tatsächlichen Sinngehalt des (vermeintlich) vermittelten Sachverhalts. So würden Feedbacks zur Lehreffektivität stärker vom Auftreten als vom tatsächlichen Lerninhalt beeinflusst. Da angesichts dieses Befunds die Zufriedenheit als durch persönliche Eigenschaften leicht zu beeinflussende und somit nicht aussagekräftige Variable anzusehen ist, die nicht in Zusammenhang mit dem Lerninhalt oder der Lehreffektivität stehen muss, zogen die Forscher den Schluss, dass es anderer Methoden bedarf, um in Ratings die Lehreffektivität zu erheben (Naftulin/Ware/Donnelly 1973).

Gemäß den Forschungen von Hofer (1981) sind Schülereinschätzungen von persönlichen Eigenschaften ihrer Lehrerinnen und Lehrer grundsätzlich durch den Dr.Fox-Effekt verzerrt. Auch Murray und Lawrence (1980) konnten nachweisen, dass Lehrkräfte, die auf bestimmte Verhaltensweisen trainiert wurden wie bspw. stimmliche Variationen und ausdrucksvolle Bewegungen, bessere Ratings erhielten als eine Kontrollgruppe. Diesen Befunden wird allerdings das Argument gegenübergestellt, dass sich diese Täuschung nicht über eine einmalige Lehrsituation hinaus aufrechterhalten lässt, da die Schülerinnen und Schüler im Schulalltag nicht dauerhaft unter dem blenderischen Einfluss verweilen, sondern ihre

Lehrkräfte relativ gut im Laufe des Schuljahres in vielen Stunden gemeinsamen Unterrichts kennenlernen. Daher spielen in dieser Lehrer-Schüler-Beziehung eher sozialisationsbedingte Erwartungsmuster und individuelle Beurteilungstendenzen eine Rolle (Ditton 2002a).

2.2.3 Lehrer aus Schülersicht

Nach Asendorpf und Neyer (2012) sind

„Beurteilungen durch Interaktionspartner (statt durch Beobachter) [sind] nur dann sinnvoll, wenn es um den subjektiven Eindruck von der Versuchsperson speziell bei Interaktionspartnern geht [...]“ (Asendorpf/Neyer 2012, S. 101)

Vor diesem Hintergrund wurden persönliche Merkmale wie Humor, Enthusiasmus oder Geduld der Lehrkraft sowie Sympathie mit der Lehrkraft oder allgemein gehaltene Effektivitätseinschätzungen (v.a. in der amerikanischen Forschung, siehe hierzu mehr in Kap. 4.3) aus Schülerperspektive erhoben. Diese können mit ihrem spezifischen Fokus auf separate Verhaltensweisen (und außerdem begrenzten und selektiven Stichproben) jedoch kaum umfassende Erkenntnisse bezüglich des Einflusses der Lehrerpersönlichkeit bieten; zumal die erhobenen Eigenschaften im engeren Sinne nicht als Persönlichkeitsmerkmale gelten (siehe Kap. 3).

Im Zuge uneinheitlicher Ergebnisse und der insgesamt lückenhaften Forschungslage zur Lehrerpersönlichkeit, die lange in der Diskussion als möglicher Verzerrungsfaktor von Schüleraussagen zur Unterrichtsqualität stand³², wiesen Jones et al. (1985) nach, dass Schülerinnen und Schüler zwei Aspekte differenzieren und mit guten Lehrkräften assoziieren: a) technische, in Form von Fachwissen und Prozessmerkmalen des Unterrichts und b) persönliche, in Form von Persönlichkeitsmerkmalen. Daraufhin versuchte Jones (1989), ein Evaluationsinstrument zu finden, das den Einfluss der Lehrerpersönlichkeit minimiert. Dabei folgte er dem Erhebungskonzept von Scriven (1981), der zuerst nach der Lehrperson und dem Unterrichtsgegenstand fragt und erst danach die Unterrichtsqualität evaluieren lässt. Entsprechend beurteilten ca. 90 College-Studierende zu zwei Erhe-

³² Bspw. Small, Hollenbeck und Haley (1982) bezeichneten Studierendenevaluationen sogar als invalide, da sie sich auf den persönlichen Kontext reduzieren. Zudem rufe Charisma den Dr.Fox-Effekt hervor, der sich in den Evaluationen spiegele (vgl. Jones et al. 1985, mehr dazu in Kap. 2.2.1).

bungszeitpunkten die Biologie-Kurse von 15 Lehrkräften (wobei nur fünf Lehrkräfte zu beiden Erhebungszeitpunkten evaluiert und zudem die Klassen von Lehrergruppen unterrichtet wurden) anhand von Einzelitems: erst zu den Konstrukten Fachinhalt, Sympathie mit der Lehrperson – was in dieser Untersuchung als Äquivalent zur Lehrerpersönlichkeitseinschätzung gilt – und Unterrichtsqualität der Lehrperson (was hier mit Lehrqualität begrifflich gleichgesetzt wurde); im weiteren dann zu Wissen, Kommunikation, Enthusiasmus, Erklärkompetenz, Struktur, Interesse und Schülerbefinden, die als Gesamtbild der Unterrichtsqualität aufgefasst werden. Im Ergebnis sind durchweg alle Korrelationen signifikant, zum zweiten Erhebungszeitpunkt fallen sie sogar noch höher aus. Besonders hoch korrelieren die Persönlichkeitsitems mit dem Globalurteil zur Unterrichtsqualität, mit dem Interesse und dem Enthusiasmus und etwas geringer mit dem Unterrichtsgegenstand. Schüleraussagen zur Unterrichtsqualität bleiben demnach signifikant verbunden mit der Lehrerpersönlichkeit, was jedoch – so argumentiert Jones (1989) – die Validität der Evaluationen nicht beeinträchtigt³³. Des Weiteren laden die Lehrerpersönlichkeit und die Lehrerkompetenz auf demselben Faktor, d.h. diese Konstrukte scheinen für (ältere) Schülerinnen und Schüler nicht voneinander unterscheidbar. Lublin und Barrand (1987) wiesen im Anschluss daran nach, dass Schülerinnen und Schüler selbst die Persönlichkeit ihrer Lehrkraft als den ihr Lernen am wesentlichsten beeinflussenden Faktor ansehen. Ditton (2002a) hingegen ist davon überzeugt, dass in Bezug auf Unterrichtsqualität keineswegs von einer Lehrerpersönlichkeit als deren Universallösung ausgegangen werden kann, ebenso wenig wie von einem diesbezüglichen *Unterrichtsrezept*. Zwar konnte er nachweisen, dass mit dem Alter der befragten Schülerinnen und Schüler die Bedeutsamkeit der personalen Beziehung zur Lehrkraft wächst, allerdings ohne diesen Befund in einen konkreten Zusammenhang mit der Wahrnehmung der Unterrichtsqualität zu stellen.

³³ Jones' (1989) diesbezügliche Argumentation basiert auf zwei konträren Grundannahmen: demnach sind entweder a) Persönlichkeitseigenschaften von Lehrkräften irrelevant für Schülereinschätzungen des Unterrichts, die demzufolge schlicht invalide sind; oder b) Schülereinschätzungen des Unterrichts sind dann valide, wenn die Schülereinschätzungen der Lehrerpersönlichkeit innerhalb eines validen Modells mit der Lehrkompetenz verknüpft sind. Auf Grundlage dieser beiden Annahmen spricht sich Jones für letztere Annahme aus, bleibt jedoch eine empirische Prüfung schuldig.

Clayson und Sheffet (2006) gingen nach der Feststellung, dass es überraschend wenige veröffentlichte Befunde hinsichtlich der Validität³⁴ von Schüleraussagen unter Beachtung der Wahrnehmung der Lehrerpersönlichkeit gibt³⁵, der Frage nach, inwiefern die Wahrnehmung der Lehrerpersönlichkeit die Unterrichtsevaluation beeinflusst. In einer Stichprobe von neun Lehrpersonen, die durch 481 Studierende einer Wirtschaftshochschule in den Bereichen Organisationsmanagement und Marketing beurteilt wurden, erhoben sie zu vier Zeitpunkten die folgenden Konstrukte: *Persönlichkeit* mittels fünf BigFive-Einzelitems³⁶, *Unterrichtsevaluation* mittels fünf Einzelaussagen zum Unterricht³⁷ (Lernatmosphäre, Erklärkompetenz, Interesse am Lernen der Studierenden, hohe Standards, Zufriedenheit mit dem Unterricht). Die Items wurden zu den übergeordneten Faktoren *Persönlichkeit* und *Unterrichtsevaluation* zusammengefasst. Im Ergebnis zeigt sich, dass die Persönlichkeitsdimensionen, außer *Extraversion*, signifikante Regressionskoeffizienten in Bezug auf *Evaluation* aufweisen, also starke Prädiktoren sind. Zudem variiert die *Unterrichtsevaluation* mit Veränderungen der Wahrnehmung der Lehrerpersönlichkeit, dies wiederum v.a. hinsichtlich der Wahrnehmung von *Extraversion*. Korrelationen zwischen übergeordnetem Persönlichkeitsfaktor und Unterrichtsevaluation nehmen über den Zeitraum der vier Erhebungen darüber hinaus immer stärker zu, wobei sie durchweg signifikant sind. Kritikpunkte dieser Studie sind allerdings die kleine Stichprobe sowie die sehr reduzierte Erfassung der Konstrukte anhand von Einzelitems, die weder differenzierte und valide Ergebnisse darstellen noch eindeutige Interpretationen zulassen.

Dem Angebot-Nutzungs-Modell von Helmke (2007) ist zu entnehmen, dass auf dem aktuellen Forschungsstand eher davon ausgegangen wird, dass die Schülerinnen und Schüler Aspekte der Lehrerpersönlichkeit nur indirekt über deren unterrichtliches Handeln wahrnehmen. Die direkte Wahrnehmung der Lehrerpersönlichkeit durch die Schülerinnen und Schüler ist mit Verfahren aus der Unterrichtsforschung zur Erfassung von Schülerperspektiven weder umfassend noch

³⁴ Schülerevaluationen waren lange Zeit als invalide (u.a. wegen ihrer Verzerrung durch die Lehrerpersönlichkeit) in Verruf geraten (vgl. Wolfer/Johnson 2003).

³⁵ Dies wurde bereits 1981 von Seiffke-Krenke (vgl. Saldern 1987) und 1998 von Wachtel konstatiert.

³⁶ Siebenstufige Likertskalierung.

³⁷ Sechsstufige Likertskalierung.

strukturiert erfasst noch systematisch mit Prozessmerkmalen der Unterrichtsqualität verbunden worden.

Dennoch gibt es Forschungsbefunde zu direkten Effekten personaler Lehrkraftmerkmale, die sich zudem allein auf die Schülerperspektiven richten: So zeigen sich Zusammenhänge von Unterrichtshandeln mit Schülereinschätzung persönlicher Eigenschaften ihrer Lehrkräfte in früheren Forschungen zumindest hinsichtlich Einfluss/Vormachtstellung/Vorherrschaft (ascendency), Verantwortung, emotionale Stabilität, Geselligkeit, originelles Denken, persönliche Beziehungen, Vitalität (Costin/Grush 1973), Dynamik, Pragmatismus, Freundlichkeit, hohe Intelligenz (Sherman/Blackburn 1975), Zugehörigkeit, Erfolg/Leistung, Ausdauer, Förderung, Verständnis, Wandelbarkeit, Extraversion (Tomasco 1980), Ausdruckskraft, Empathie, Strukturiertheit, Beispielverwendung, Individualisierung (Mintzes 1979), Enthusiasmus, Interaktion (Solomon 1966), Strukturierung des Materials, geschulte Präsentation, Fachwissen (Deshpande/Webb/Marks 1970), Rücksichtnahme, Kompetenz (Keaveny/McGann 1978). Inwiefern diese Eigenschaften systematisch eingeordnet werden und als zeitlich einigermaßen stabile Persönlichkeitsmerkmale gelten, lässt sich allerdings nicht ausmachen, zumal ihnen kein in sich kohärentes Konstrukt wie jenes der BigFive (vgl. Kap. 3.3) zugrunde liegt.

Auch neueste Forschungen wie die von Neuber, Künsting und Lipowsky, vorgestellt auf der Tagung der DGfE 2014 in Berlin (eingebettet in das Projekt „Wege in den Beruf“), diskutieren auf einer neu aufgestellten Argumentationsbasis die Notwendigkeit der Berücksichtigung von Schülerperspektiven auf persönliche Aspekte von Lehrkräften – allerdings lediglich auf den Aspekt des Lehrerenthusiasmus – in Zusammenhang mit der Unterrichtsqualität. Zusätzlich werden Fächervergleiche im Sinne differentieller Zusammenhänge der Fächer Mathematik und Deutsch angestellt. Noch ist die einschlägige Befundlage allerdings sehr schwach ausgeprägt und lässt kaum Einblicke in Zusammenhänge von Schülerperspektiven auf die Lehrerpersönlichkeit und die Unterrichtsqualität zu. Auch Kunter et al. (2008) zeigen Ansätze zur Erforschung von Enthusiasmus und Unterrichtsqualität. Erschwerend für eine interpretative Ableitung der Befunde auf die Lehrerpersönlichkeit ist jedoch v.a. der Umstand, dass Enthusiasmus im engeren Sinne nicht als Persönlichkeitsmerkmal gilt (vgl. Kap. 3).

2.2.4 Lehrerpersönlichkeit – Selbsteinschätzungen vs. Fremdeinschätzungen

Es lässt sich bislang genau eine Studie finden, die systematisch den Zusammenhang zwischen Selbst- und Schülereinschätzungen der Lehrerpersönlichkeit entlang der BigFive³⁸ untersucht: Mayr und Neuweg (2006) stellen dabei die mit dem ausführlichen NEO-PI-R erfasste Selbsteinschätzung der Lehrkräfte (N=60) den Schülereinschätzungen (N≈1300) gegenüber (siehe Tab. 1). Die Version der Selbsteinschätzung gründet dabei auf 240 Items, jene der Schüler(fremd)einschätzungen dagegen auf 120 Items. D.h. trotz Konstruktidentität unterscheiden sich die Instrumente. Zusätzlich wurde im Rahmen der Untersuchung die Lehrersicht auf das eigene Führungsverhalten sowie die unterrichtsmethodische Grundeinstellung erfasst. Die durchschnittlichen Reliabilitäten der Persönlichkeitsdimensionen liegen bei den Schülerinnen und Schülern bei 0,87-0,97, bei den Lehrerinnen und Lehrern etwas niedriger bei 0,87-0,91. Die höhere Reliabilität bei den Schülerinnen und Schülern erklärt sich womöglich durch deren Aggregation auf Klassenebene. Die berichteten Korrelationen zwischen Selbsteinschätzung und Schülereinschätzungen weisen folgende Stärken innerhalb der Dimensionen auf:

Tab. 1: Korrelationen zwischen Selbsteinschätzung und aggregierten Schülereinschätzungen (N=60, Instrument: NEO-PI-R, Signifikanz: $p > 0,01$) (nach Mayr/Neuweg 2006, Signifikanzen ergänzt durch Mayr 2019) – Die Korrelationen innerhalb der Dimensionen Offenheit und Extraversion liegen leicht unterhalb derer von Selbst- und Bekanntheitschätzung (die bei Offenheit mit 0,61 und bei Extraversion mit 0,53 korrelieren).

Gewissenhaftigkeit	Extraversion	Offenheit	Neurotizismus	Verträglichkeit
,19	,49**	,41**	,18	,11

Zudem zeigen sich dimensionsübergreifende signifikante Zusammenhänge ($p > 0,01$) zwischen *Extraversion* und *Offenheit* (Mayr 2019). Offenbar wird trotz identischer Itemformulierungen unterschiedliches gemessen oder aber Schülerinnen und Schüler schätzen ihre Lehrerinnen und Lehrer tatsächlich anders ein als diese sich selbst (Mayr/Neuweg 2006). Die Autoren schließen hieraus, dass auf-

³⁸ Es ist der Stringenz der Arbeit geschuldet, dass das Konstrukt der BigFive erst in Kapitel 3 präzise erläutert wird.

grund der breiteren Erfahrungsgrundlage der Lehrkräfte deren Einschätzungen valider sind als die Schülereinschätzungen. Angesichts der theoretischen Überlegungen zum Konstrukt Persönlichkeit (siehe Kap. 3) in Zusammenhang mit der Perspektive (Kap. 1.1) ist diese Interpretation zumindest zu hinterfragen.

Wird die unterrichtsmethodische Grundeinstellung der Lehrkraft in die Korrelationen als Kovariate einbezogen, sinken bei Lehrkräften mit einer konstruktivistischen Orientierung die Korrelationen zwischen Lehrerselbst- und Schülerfremdeinschätzungen in den Dimensionen *Extraversion* und *Offenheit für neue Erfahrungen* auf 0,40 (*Extraversion*) und 0,32 (*Offenheit für neue Erfahrungen*) (Mayr/Neuweg 2006).

2.2.5 Assessment-Center

Der Einsatz von Persönlichkeitsfaktoren in der Personalauswahl sowie innerhalb von Betriebsevaluationen ist heute nicht mehr unüblich. Dazu werden zunächst Eigenschaften von Mitarbeitern gemessen und mittels linearer Beziehungen und/oder Korrelationen in Verbindung mit deren Berufserfolg gebracht. Daraus leiten Assessment Center arbeitsbezogene Fähigkeiten und soziale Kompetenzen ab, die bei der Personalauswahl durch mehr oder weniger aufwendige Ermittlung Beachtung finden (Asendorpf/Neyer 2012). Die START-P-Testbatterie (Beaucucel/Kersting 2010) ist z.B. ein Online-Berufseignungstest zur Hilfe bei berufsrelevanten Entscheidungen. Der Fokus richtet sich dabei auf Jugendliche und junge Erwachsene. Zur Bestimmung des Berufserfolgs wird u.a. anhand der BigFive (s. Kap. 3.3) die selbsteingeschätzte Persönlichkeit erfragt (vertiefend siehe Wenzel 2011). Die Testbatterie unterscheidet hier bspw. zwischen verhaltensinput- und verhaltensoutputbezogenen Persönlichkeitseigenschaften bezogen auf Reiz (=Verhaltensinput) und die äquivalente Reaktion (=Verhaltensoutput) nach Allport (1961). Der Verhaltensinput bildet dabei die Reizsensitivität ab, die interindividuell variiert und zu einem entsprechenden Verhaltensoutput in Form von persönlichen Merkmalen wie Gewissenhaftigkeit, Resilienz oder Extraversion führt, die ihrerseits als Prädiktoren für die berufliche Eignung gelten (Kersting 2005, Hanfstingel/Mayr 2007, Wenzel 2011).

In den 1960er bis 1980er Jahren herrschten pessimistische Einstellungen gegenüber dem Nutzen von Persönlichkeitserhebungen für die Vorhersage beruflicher

Leistungen³⁹. Das änderte sich seit den 1990er Jahren mit der Publikation einschlägiger Metaanalysen (Barrick/Mount 1991, Barrick/Mount/Judge 2001, Hurtz/Donovan 2000), die Persönlichkeitsskalen durchaus als valide Prädiktoren für die berufliche Leistung auswiesen (Hülshager/Specht/Spinath 2006). Zudem wurde in den darauffolgenden Jahren die Lehreffektivität verstärkt als Teil der beruflichen Leistung von Lehrerinnen und Lehrern und deren Persönlichkeit somit als Prädiktor sowohl für Lehreffizienz als auch für die Berufsleistung angesehen. Vor allem das BigFive-Modell (vgl. Kap. 3.3) gilt als prädiktiv valide hinsichtlich der Berufseinstellung (Salgado 1997), obwohl die meisten Studien die Persönlichkeit nicht entlang der BigFive und selten überhaupt systematisch abbilden.⁴⁰ Nachdem das Thema und die Erhebung von Lehrerpersönlichkeiten jedoch angesichts indifferenter Befunde mehrere Jahre in Verruf geraten waren, fand es angesichts des Eingangs von Persönlichkeitseignungstests im wirtschaftlichen Personalmanagement auch im Bildungswesen wieder Einzug. Es herrschte zudem weitgehend Einigkeit innerhalb der empirischen Persönlichkeitsforschung, dass Erhebungsverfahren entlang der BigFive in vielerlei Hinsicht als zweckdienlich hinreichend valide gelten können, womit sich Persönlichkeit systematisch anhand ihrer verschiedenen Dimensionen und Facetten erheben ließ (Mayr 2003). Bereits 1984 verknüpfte Urban Persönlichkeit (Selbsteinschätzung) mit Unterrichtskompetenz (Beobachter) mit dem Ergebnis, dass sich Zusammenhänge zwischen Persönlichkeit sowie Handeln und Befinden zeigten (Urban 1984). Ergebnisse weiterer Studien als auch Metaanalysen (Tokar/Fischer/Subich 1998, Judge/Heller/Mount 2002) stellten in den 1990ern in Bezug auf das berufliche Befinden und die berufliche Leistung die Dimensionen *Neurotizismus* (resp. *Resilienz*), *Extraversion* und *Gewissenhaftigkeit* heraus; dieselben Befunde ließen sich für Lehrkräfte insofern bestätigen (insbes. von Mayr 1994,

³⁹ Die prädiktive Validität ($M=0,37$, $SD=0,13$) wurde angesichts variierender Ergebnisse eher als gering und wenig aussagekräftig eingeschätzt (Asendorpf/Neyer 2012).

⁴⁰ Es macht eher den Anschein, dass die Schulforschung hier Untersuchungen und Metaanalysen über alle Berufsgruppen hinweg folgte, wonach die Dimension *Gewissenhaftigkeit* (Barrick/Mount 1991), neben *Intelligenz*, *Offenheit für neue Erfahrungen* und *Verträglichkeit* in allen Berufen die Berufsleistung valide vorhersagt. Es erscheint allerdings fraglich, ob vor allem der objektive Berufserfolg der untersuchten Berufsgruppen, die sich überwiegend auf den Finanzdienstleistungssektor beschränken, mit dem Berufserfolg von Lehrerinnen und Lehrern im Sinne von Lernerfolgen/Lernleistungen ihrer Schülerinnen und Schüler vergleichbar ist.

2003), als die BigFive engagiertes Studierverhalten, Studienleistungen, Kompetenz und Wohlbefinden im Beruf prognostizieren. Das Lernen im (Lehramts-) Studium (in Form von: Nutzung der Angebote, Anwendung von Lernstrategien und Ausdauer) ist demnach nachweislich beeinflusst durch personale Voraussetzungen: im Überblick über mögliche, persönliche Faktoren als Prädiktoren für den Studienerfolg ragen die Merkmale *Extraversion*, *Resilienz*, *Gewissenhaftigkeit* stets heraus. So ist ein Mindestmaß an *Extraversion*, *Resilienz*, *Gewissenhaftigkeit* nötig für ein erfolgreich abgeschlossenes Lehramtsstudium. Diese Persönlichkeitsmerkmale im Sinne von Persönlichkeit „als Ensemble (relativ) stabiler Dispositionen, die für das Handeln, den Erfolg und das Befinden im Lehrerberuf bedeutsam sind“ (Mayr/Neuweg 2006, S.183), wurden zudem als prädiktiv für den Studienerfolg im Längsschnitt belegt (Mayr 2014) und gelten seitdem als Voraussetzung für das eigene Lernen von (angehenden) Lehrkräften (siehe Baumert/Kunter 2010, Lipowsky 2010, Mayr 2014). Mayr nennt die „Persönlichkeit eine wichtige Determinante der pädagogischen Kompetenz“ (Mayr 2003, S. 90) und in diesem Zusammenhang die Notwendigkeit persönlichkeitsfördernder Bildungsangebote, wobei unterschiedliche Voraussetzungen einer bunten Vielfalt an Angeboten bedürfen. Darüber hinaus legt Mayr nahe, Persönlichkeitsmerkmale bei der individuellen Entscheidung für oder gegen ein Lehramtsstudium zu berücksichtigen (Mayr 2000, 2003).

Dieser Erkenntnis entspricht auch der Befund, dass objektiver und subjektiver Berufserfolg in positiver Beziehung zueinander stehen (positiv korrelieren) und sich für beide v.a. die Dimensionen *Neurotizismus* (resp. *Resilienz*), *Gewissenhaftigkeit* und *Extraversion* prädiktiv relevant erweisen (Hülshager/Specht/Spinath 2006). Allerdings zeigen andere Befunde, dass nur *Neurotizismus* ein starker Prädiktor sowohl für objektiven als auch subjektiven Berufserfolg bzw. -misserfolg ist (Boudreau/Boswell/Judge 2001, Judge/Heller/Mount 2002, Barrick et al. 2001), während sich *Gewissenhaftigkeit* in neueren Studien als nicht signifikant zeigt (Boudreau et al. 2001 in Hülshager/Specht/Spinath 2006). Bipp (2007) untersuchte darüber hinaus Auswirkungen der BigFive-Persönlichkeitsmerkmale auf Handlungsprozesse: Ausgangspunkt der Untersuchung ist, dass über verschiedene Berufsgruppen hinweg vor allem *Resilienz* und *Gewissenhaftigkeit* eine Prädiktionskraft bzgl. beruflicher Leistungsmaße aufweisen (Barrick/Mount/Judge 2001). Die Studie von Bipp (2007) weist darüber hinaus eine Wirkung dieser beiden Dimensionen auch in Bezug auf Handlungsprozesse bei einfachen Leistungsaufgaben nach, d.h. Eigenschaften gehen vermittelt über den Handlungsprozess in die

Arbeitsleistung ein. Die Prädiktionskraft variiert dabei nicht zwischen den Subfacetten und Globalfaktoren, allerdings besitzen die Subfacetten höhere Erklärungswerte (Bipp 2007).

Mit diesen Befunden geht allerdings die Desillusionierung von Vorstellungen einher, die auf der Lernbarkeit von Unterrichts- und Erziehungskompetenz basieren (Mayr/Neuweg 2006). Laut Hanfstingel und Mayr (2007) lauern in Bezug auf die Prognose der Bewährung im Lehramtsstudium daher zwei Gefahren: zum einen die der *Self-fulfilling Prophecy*, womit die freie Entfaltung gehemmt würde; und zum anderen die der politischen Steuerung, wobei der Zugang zur Lehrerausbildung selektiv kontrolliert und damit einhergehend das aktuell politische System stabilisiert würde. Außerdem sehen Mayr und Neuweg (2006) in einer ausgeprägten Stabilitätsannahme von Persönlichkeit die Gefahr der Selektion zwischen Geeigneten und Ungeeigneten ohne Beachtung des möglichen Entwicklungspotenzials. Allerdings sehen sie ebenso eine Gefahr in der Annahme einer veränderbaren Persönlichkeit: Sie könnte unter der Auffassung der beliebigen Modifizierbarkeit von Persönlichkeitsmerkmalen dann nicht Eingangskriterium, sondern Zielgröße darstellen (vgl. Ideen von der Persönlichkeitsbildung durch Selbstentwicklung (Schulz von Thun 2005), durch persönlichkeitsfördernde Veranstaltungen oder durch Gestaltung der institutionellen Umwelt (Teml/Unterweger 2002)). Es ließen sich demgegenüber aber auch Chancen im Sinne der Formulierung von Entwicklungszielen und der Vervielfältigung von Lernarrangements ausmachen (Nieskens 2002), je nachdem, ob deterministischen oder probabilistischen Annahmen gefolgt wird. Der Mittelweg sähe nach Mayr (und Neuweg 2006) demnach so aus, dass – ohne Persönlichkeitsveränderungen auszuschließen – der Fokus auf den bewussten Umgang mit Persönlichkeit durch individuelle Selbsterkenntnis (Umgang mit Stärken und Schwächen) gelegt würde, in Form personenbezogener Reflexion, institutioneller Differenzierung und Individualisierung der Lernangebote.

Zur Selektion über die Persönlichkeitsbeurteilung durch externe Personen in Assessment Centers befinden Mayr und Neuweg (2006) folgendermaßen:

„Derartigen Maßnahmen kann man mit guten Gründen das Argument einer mangelnden diagnostischen und prognostischen Validität nicht nur der intuitiven sondern auch der objektiven Verfahren der Persönlichkeitsbeurteilung und einer insgesamt doch zu geringen Varianzaufklärung entgegenstellen. Wir teilen diese Bedenken ausdrücklich.“ (Mayr/Neuweg 2006, S. 201)

Dennoch zeigen sich mittlerweile auch in der Lehrerausbildung vermehrt Ansätze (bspw. PArCour der Universität Passau⁴¹, TESAT in Österreich (Neubauer/Koschmieder/Krammer/Mayr et al. 2017)), die Persönlichkeitsinventare in ihre Selektionsverfahren aufzunehmen gedenken – ausnahmslos aus Perspektive des Selbstberichts heraus –, um geeignete Kandidaten von weniger oder ungeeigneten zu unterscheiden. Hierbei muss festgehalten werden, dass die Facetten der BigFive (Poropat 2009) und Leistungs- und Fähigkeitstests zwar durchaus prädiktiv valide sind in der Hinsicht, als dass sie sich zur Vorhersage des Studienerfolgs eignen; nicht kognitive Persönlichkeitsmerkmale eignen sich zudem zur Vorhersage der Studienzufriedenheit (Rindermann/Oubaid 1999). Gut erforscht sind vor allem positive Zusammenhänge von *Gewissenhaftigkeit*, *Extraversion* und *Neurotizismus* mit dem Studienerfolg (Mayr 2014). Profile von Lehrkräften und Lehramtsstudierenden weisen zudem sehr hohe Ähnlichkeiten untereinander auf (Schaarschmidt/Fischer 2013). Demnach gelten Persönlichkeitsmerkmale als Mehrgewinn für Zulassungsverfahren (Kuncel/Hezlett, 2010). Mayr (2014) zieht darüber hinaus den Schluss, dass sich die in Bezug auf den Studienerfolg herausgestellten Dimensionen *Gewissenhaftigkeit*, *Extraversion* und *Neurotizismus* ebenso in Praxisleistungen resp. der Unterrichtsgestaltung niederschlagen und demzufolge bei der Selektion von Lehramtsanwärtern von besonderer Bedeutung sind. Inwiefern sich diese Dimensionen in Praxisleistungen von Lehrerinnen und Lehrern im Dienst niederschlagen, ist aber nicht nachweisbar, zumindest nicht, wenn die entsprechende Praxisleistung als hohe Unterrichtsqualität definiert wird. In der Studie von Mayr und Neuweg aus dem Jahr 2018 zeichnen sich zwar eher schwache Zusammenhänge zwischen der unterrichtlichen Grundeinstellung von Lehrkräfte und Ihrer selbsteingeschätzten Persönlichkeit ab; allerdings ebenso mit dem Alter und der Dauer der Berufspraxis. Die Autoren leiten daraus die Vermutung ab, „dass dem hochschuldidaktischen Bemühen um innovativen Unterricht durch relativ stabile Persönlichkeitsmerkmale Grenzen gesetzt sein könnten“ (Mayr/Neuweg 2018, S. 10). Vor diesem Hintergrund erscheint die Selektion angehender Lehrkräfte entlang ihrer (selbstwahrgenommenen) Persönlichkeit über ökonomische Überlegungen, die sich auf den Studienerfolg oder die

⁴¹ „Derzeit ist PArCours ein Beratungsangebot. Sie behalten unabhängig von den Ergebnissen in jedem Fall Ihren Studienplatz.“, Näheres unter <http://www.phil.uni-passau.de/schulpaedagogik/forschungprofilelemente/parcours/ueber-parcours/>, siehe auch Hechinger/Seibert 2015.

berufliche Zufriedenheit beziehen, hinaus als wenig fundiert. Eher erinnert es an Modelle der Hypernormalisierung (siehe bspw. Maaß 2014, Cuntz/Krause 2012, zur weiteren Vertiefung siehe Link 2006), in deren Konsequenz letztlich eine Einschränkung von Pluralität zu befürchten wäre.

3 Dialektik der Persönlichkeit – Schnittstelle zwischen Innen- und Außenwelt

Eine Persönlichkeit hat jeder Mensch, ihr Dasein ist selbstverständlich und allgegenwärtig. Ein allgemeines, einheitliches und klar definiertes Begriffsverständnis von ihr scheint jedoch nicht zu existieren, da sie kaum mit Worten greifbar ist (Backhaus 2004). Abzuleiten ist der Begriff *Persönlichkeit* von seinem lateinischen Wortstamm *persona*, wie die Theatermasken von Schauspielern bezeichnet wurden. Diese Masken dienten dazu, relevante Eigenschaften der dargestellten Charaktere zu veranschaulichen, hervorzuheben, widerzuspiegeln. Entsprechend der Bedeutung des erweiterten lateinischen Wortstamms *personare* und *persono*, die sinngemäß mit *hindurchtönen*, *laut rufen* oder *widerhallen* übersetzt werden (Langenscheidts Universal-Wörterbuch 1994), offenbarte sich die Rolle der gespielten Person verstärkt durch die zur Schau gestellte Maske und wurde für die Zuschauer berechenbar und nachvollziehbar. Dank der *persona* war unmissverständlich absehbar, wie die Figur in bestimmten Situationen reagieren würde (Backhaus 2004, Rammsayer/Weber 2010). Angesichts seiner ursprünglichen Semantik wird deutlich, dass sich *Persönlichkeit* vor allem auf das nach außen und öffentlich dargebotene Bild eines Menschen bezog, das sich in Form bestimmter und von anderen Menschen wahrnehmbarer Verhaltensweisen manifestiert (Rammsayer/Weber 2010).

Mit der Etablierung von (modernen) Vorstellungen des Menschen als eines Wesens, das sich sowohl aus für andere sichtbaren als auch von außen nicht einsehbaren Anteilen zusammensetzt (aufbauend auf Theorien von Mead⁴², Goffman⁴³,

⁴² Sowohl das gesellschaftliche *Me* als auch das spontan reagierende *I* sind nach Mead (1934) erforderlich für eine Identität. Da das *I* durch Spontaneität definiert ist, ist es lediglich nachträglich erfassbar. Durch diese Nachträglichkeit wird es jedoch zu einer reflektierten, auf einer Perspektivenübernahme beruhenden Version des *Me*. Erst die Einheit von Perspektivenübernahmen (*Me*) und Reaktion (*I*) bilden nach Mead das Selbst, das situationsabhängig, spontan und nicht unveränderlich fest ist. Indem das Ich bei Mead ausschließlich als Resonanz auf andere existiert und somit gesellschaftlich vermittelt ist, wird ein Ich als fester Identitätskern bestritten (mehr hierzu siehe bei Peatow 2004 und Schäfer 2005).

⁴³ Nach Goffman (1975) steht die Persönlichkeit aufgrund der Abhängigkeit von der wechselseitigen Anerkennung in allen zwischenmenschlichen Interaktionen auf dem Spiel. In

Habermas⁴⁴ etc.), erschien das Verständnis von *Persönlichkeit* als ein Konstrukt, das ausschließlich auf durch andere Personen wahrnehmbare Eigenschaften basiert, verkürzt. Seither werden unter dem Aspekt *Persönlichkeit* auch solche Anteile des Menschen gefasst, die im Verborgenen und für andere nicht direkt sichtbar wirken. Auffällig an der weiteren Entwicklung der Wortbedeutung von *Persönlichkeit* ist, dass diese zunächst auf äußerlich sichtbare Aspekte fokussierte Definition (äquivalent zur Entwicklung von Subjekttheorien) immer weiter von außen nach innen verlagert wurde. Innerhalb eines subjektoptimistischen Neuhumanismus⁴ findet sie sich letztlich als identische Einheit innerer und äußerer Merkmale zusammengefasst wieder, was jedoch offenbar vielmehr ein Spannungsfeld zwischen einem kontinuierlichen, harmonischen Fortschreiten hin zur Selbstvervollkommnung und autonomen, widerständigen Anteilen des Selbst abbildet (vgl. Adorno in Meyer-Drawe 2000). Trotz der Bedeutung sowie den Unvereinbarkeiten von inneren und äußeren Aspekten lesen sich Definitionen zur Persönlichkeit in der neueren Forschungsliteratur durchweg ähnlich wie die folgende:

„Mit Persönlichkeit wird das einem Menschen spezifisch organisierte Gefüge von Merkmalen, Eigenschaften, Einstellungen und Handlungskompetenzen bezeichnet.“ (Hurrelmann 1993, S. 14)

Asendorpf hingegen bezeichnet Persönlichkeit erst als „stabile Beziehungen zwischen den Situationen und den Reaktionen einer Person“ (2004, S. 36), dann – mit einem modifizierten Fokus des Subjektverständnisses – als die „Gesamtheit aller überdauernden individuellen Besonderheiten im Erleben und Verhalten eines Menschen“ (in Wirtz 2014, S. 1244), wobei unter *überdauernd* ein Zeitraum von

Interaktionen implizieren die Erwartung der Anderen, dass gemeinsame soziale Regeln gültig sind und sich dadurch das *Ich* offenbart; dabei verschwimmt die Trennung zwischen Sein und Schein. Da Interaktion und Kommunikation auf wechselseitiger Interpretation beruhen und keine Sicherheiten liefern, kann keine eindeutige, gefestigte Persönlichkeit erlangt werden (mehr hierzu siehe bei Schäfer 2005).

⁴⁴ Habermas (1976) definiert *Persönlichkeit* entlang verschiedener Identitäten, bspw. *persönlcher* und *sozialer Identität*, die zwar miteinander verknüpft aber nicht dasselbe sind. Unter *Ich-Stärke* fasst er dabei die persönliche Stabilität innerhalb widersprüchlicher Diskurse auf, die auf der Unwiderlegbarkeit rationaler Argumentation gründet und das *Ich als Herr im eigenen Haus* zum Ziel hat (mehr dazu bei Paetow 2004, Schäfer 2005, Meyer-Drawe 2000, Altwicker 1999).

einigen Wochen oder Monaten, also eine im Vergleich zur gesamten Lebensspanne eher kurzfristige Stabilität zu verstehen ist; Backhaus (2004) hingegen entwirft ein Modell, in dem Persönlichkeit als die Summe aller Erbanlagen, erlernten Bewertungen und Reaktionen erscheint. Diesen Begriffsinterpretationen steht die frühe Interpretation Allports (1937) entgegen, der die Definition von *Persönlichkeit* weitgehend fluid, d.h. dynamisch, veränderbar und veränderlich zu halten versuchte, indem er die kontextbedingte Bedeutungsvariation entlang unterschiedlichster Interpretationen und Zuschreibungen postulierte (Rammsayer/Weber 2010). So definiert er Persönlichkeit als „dynamische Organisation derjenigen Systeme im Individuum, die sein charakteristisches Verhalten und Denken determinieren“ (Allport in Backhaus 2004, S. 2) – trotz des Aspekts der Dynamik erscheinen innere und äußere Merkmale hierbei weiterhin innerhalb einer Einheitlichkeit verhaftet. Im umgangssprachlichen Gebrauch wird die *Persönlichkeit* im Allgemeinen dadurch erfasst, dass „[...] wir einen Menschen in seinem Verhalten und Erleben mehr oder weniger umfassend beschreiben“ (Rammsayer/Weber 2010, S. 12), was bei genauer Betrachtung auf die Feststellung hinausläuft, dass sich Persönlichkeit anhand von Persönlichkeitsmerkmalen resp. -eigenschaften identifizieren lässt, was auch der aktuellen Definition von *Persönlichkeit* im Duden entspricht, die diese als „Gesamtheit der persönlichen (charakteristischen, individuellen) Eigenschaften eines Menschen“ beschreibt.

Analog zum umfassenden Persönlichkeitsbegriff stellen auch Haller und Müller (2006) fest:

„Persönlichkeitsmerkmale bestimmen zugleich in starkem Maße Handeln, Einstellungen und Emotionen. Unterschiedliche Reaktionen von Personen in gleichen Situationen lassen sich auf diese Persönlichkeitsmerkmale zurückführen.“ (Haller/Müller 2006, S. 11)

Ebenso definieren Asendorpf und Neyer (2012) den Begriff der Persönlichkeit folgendermaßen:

„Persönlichkeit ist die nichtpathologische Individualität eines Menschen in körperlicher Erscheinung, Verhalten und Erleben im Vergleich zu einer Referenzpopulation von Menschen gleichen Alters und gleicher Kultur. Persönlichkeit wird von der empirischen Persönlichkeitspsychologie erforscht, die den Anspruch hat, den Kriterien empirischer Wissenschaft zu genügen.“

Insbesondere wird gefordert, dass Persönlichkeitseigenschaften operationalisiert und persönlichkeitspsychologische Aussagen empirisch prüfbar sind.“

(Asendorpf/Neyer 2012, S. 20)

Es scheint offensichtlich, dass in diesen Definitionen, die das gängige Verständnis von *Persönlichkeit* widerspiegeln, durchaus – wenn auch eher nebulös – zwischen äußeren und inneren Merkmalen differenziert wird bzw. zwischen solchen, die das Verhalten, und solchen, die das Erleben betreffen: während das Erleben kaum von außen direkt einsehbar ist, können bzgl. des Verhaltens durchaus Merkmale von anderen resp. von außen zugeschrieben werden, die wiederum Rückschlüsse auf innere Merkmale zulassen. Ausgenommen von derartigen Persönlichkeitsmerkmalen sind in der Regel jedoch physische und leistungsbezogene Merkmale ebenso wie das Temperament der entsprechend beschriebenen Person (Rammseyer/Weber 2010), was dem Befund entspricht, dass Laienurteile über Persönlichkeit generell eher eine oberflächliche Identifizierung unter Auslassung spezifischer Eigenschaften abbilden (Lang/Lüdtke 2005). Demzufolge ließe sich *Persönlichkeit* in einer ersten Annäherung an eine Eingrenzung des Begriffs über Merkmale erschließen, wobei solche Merkmale wie Intelligenz, motorische Geschicklichkeit, optische Erscheinung ebenso auszuschließen sind wie Enthusiasmus oder Optimismus (vgl. Rammseyer/Weber 2010, Lang/Lüdtke 2005).

Es wird also zwischen *Persönlichkeit* und *Persönlichkeitsmerkmalen* resp. *-eigenschaften* oder *Wesenszügen* (Wirtz 2014) unterschieden, wobei bei näherer Betrachtung Unklarheiten aufkommen, in welcher Beziehung diese Konstrukte zueinander stehen: Meist wird die Persönlichkeit als Summe, Struktur oder Muster einzelner Persönlichkeitsmerkmale bezeichnet (Wirtz 2014, Backhaus 2004). Als solche Merkmale oder Eigenschaften gelten dabei das individuelle Erscheinungsbild, das Verhalten und erst bei näherer Bekanntschaft das Erleben (Asendorpf/Neyer 2012) oder Erbanlagen, erlernte Bewertungen und Reaktionen (Asendorpf 2014)⁴⁵ oder auch die begriffliche Erfassung bestimmter Kombinationen von Handlungsweisen, die lediglich interpretative Rückschlüsse auf das innere Erleben zulassen (Backhaus 2004). In engem Zusammenhang mit Merkmalen und Eigenschaften der Persönlichkeit steht zudem der Begriff der

⁴⁵ Was wiederum dem Konzept der personalen Identität im Sinne der individuellen Erscheinung, die auch die Persönlichkeitsmerkmale einer Person enthält, entspricht und die als „Dialektik zwischen Natur und Gesellschaft“, also als Wechselwirkung aus Veranlagung und Umwelt (Haller/Müller 2006, S. 13) gedacht wird.

Disposition. Dispositionen erscheinen dabei zum einen als das sichtbare Verhalten generierende, dabei jedoch selbst nicht sichtbare Merkmale (Asendorpf/Neyer 2012), zum anderen als Synonym für Eigenschaften im Sinne von Verhaltens- und Erlebenstendenzen in spezifischen Situationen (Asendorpf 2014), in Bezug auf eine zeitliche Stabilität als Denk-, Gefühls- und Verhaltensmuster dauerhaft prägende Instanz (Iser/Schmidt 2005).

Demgegenüber lässt sich festhalten, dass insgesamt die je nach Disziplin variierenden Bezeichnungen für das Konstrukt, das in dieser Arbeit als *Persönlichkeit* bezeichnet wird, Allports frühem (1937) Postulat der Dynamik zumindest weitgehend entsprechen. Die unterschiedlichen Begriffe wie *Identität*⁴⁶, *Subjekt*, *Selbst* oder *Persönlichkeit* suggerieren, dass es sich bei ihnen um gänzlich unterschiedliche Konstrukte handelt.⁴⁷ Allerdings verhält es sich vielmehr so, dass unterschiedliche Forschungstraditionen und -paradigmen (dazu mehr im nächsten Kapitel) mit unterschiedlichen Begrifflichkeiten desselben Gegenstands arbeiten und die Verschiedenheit der Begriffe nicht die Verschiedenheit der Gegenstände an sich, sondern die Verschiedenheit der Perspektive auf und damit die Verschiedenheit der Konzeption ein und desselben Konstrukts ausdrückt (Wirtz 2014, Mayr 2003). Während in der Soziologie von *Identitätsforschung* die Rede ist, ist es in der Sozialpsychologie die *Selbstkonzeptforschung* und in der allgemeinen

⁴⁶ Laut Duden bezeichnet der Begriff *Identität* zum einen die „Echtheit einer Person oder Sache“ als „völlige Übereinstimmung mit dem, was sie ist oder als was sie bezeichnet wird“, im psychologischen Kontext zudem „als ‚Selbst‘ erlebte innere Einheit der Person“ (URL: <http://www.duden.de/rechtschreibung/Identitaet#Bedeutung1a>; Stand 24.06.2016). Demnach würde sich der Identitätsbegriff vornehmlich auf das innere Empfinden von Personen beziehen, was allerdings nicht den in entsprechenden Theorien gängigen Begriffsverwendungen entspricht (siehe vertiefend hierzu bspw. Oppenrieder/Thurmair 2003).

⁴⁷ Dieser Annahme folgen bspw. Oerter und Montada, wenn sie feststellen, dass die Begriffe *Selbst* und *Identität* zwar bedeutungsgleich verwendet, jedoch als individuelle Persönlichkeitsstruktur in Verknüpfung „mit dem Bild, das andere von dieser Persönlichkeitsstruktur haben“ (Oerter/Montada 2002, S. 11) und unter Einbezug des Selbstkonzepts verstanden werden. Allerdings fallen derartig präzierte Definitionen je nach Autor anders aus. So differenziert bspw. Mecheril (2006) die Begriffe *Subjekt* und *Identität*, indem er das Subjekt als Träger von Identität auffasst, wobei das Subjekt das *bloße* Sein darstellt und die Identität dessen *In-der-Welt-Sein* in Form seines unvermeidlichen, ständigen Ausdrucks.

Psychologie die *Persönlichkeitsforschung* oder auch die *Subjektforschung*⁴⁸. Innerhalb der Pädagogik ist die diesbezügliche begriffliche Verwendung allerdings völlig undefiniert, da hier die verschiedensten Disziplinen eingebunden werden. Meistens wird dann auf den Persönlichkeits- oder den Identitätsbegriff zurückgegriffen (gelegentlich auch auf den Subjektbegriff) oder die Semantik der jeweiligen Disziplin wird schlicht adaptiert. Eine Definition oder immerhin eine Präzision bleibt allerdings in der Regel aus: Es wird entweder vorausgesetzt, dass ein einvernehmliches Verständnis des Gegenstands bzw. des Konstrukts der *Persönlichkeit* herrscht oder – so hat es zuweilen den Anschein – die Auseinandersetzung mit der Komplexität des Konstrukts wird durch die Selbstverständlichkeit der Verwendung schlichtweg umgangen (siehe bspw. Helmke 2009, 2014, Hattie 2013). Sofern in der pädagogischen Literatur auf die Persönlichkeit bezogene Definitionen erfolgen (siehe Zitate oben), bleiben diese – ganz in pädagogischer Tradition – innerhalb eines modernen subjekt-optimistischen Weltbildes verhaftet, indem diese so gefasst werden, als seien Erleben (im Sinne von Fühlen/Empfinden und Denken) und Verhalten (im Sinne von Handeln) selbstverständlicherweise identisch (vgl. bspw. Paetow 2004). Aus diesem reduzierten Verständnis des Konstrukts *Persönlichkeit*, das von der einen wahren, in sich identischen Persönlichkeit ausgeht, ließen sich möglicherweise viele der Missverständnisse, auf denen Probleme bei der empirischen Erfassung von (z.B. Lehrer-)Persönlichkeiten gründen, ableiten (mehr dazu in den Kapiteln 2.2 und 3). Es dürfte heutzutage allerdings kaum mehr selbstverständlich vorausgesetzt werden, dass innere und äußere Welt identisch sind: die theoretischen Denkmodelle (bspw. der Konstruktivismus der Postmoderne) sprechen vielmehr für eine (oder auch mehrere) *innere* und eine (oder mehrere) *äußere* Persönlichkeit(en), die zwar in ständiger Wechselwirkung stehen, jedoch nicht als identisch oder ein und dasselbe aufgefasst werden und eine der Persönlichkeiten nicht als mehr oder weniger zutreffend als eine andere gilt. Die Grundlagen für diese dualistische Konzeption des Subjekts lassen sich zwar bereits bei Denkern der Antike sowie bei Locke, Hume und anderen finden; als deutsch-amerikanischer Vorläufer innerhalb der Moderne gilt jedoch Erikson (1966): Er ging von einem sich selbst

⁴⁸ Zwar gibt es Ansätze, die versuchen, Differenzen zwischen den Konstrukten aufzuzeigen, etwa indem *Identität* als das integrierte Ganze des Nebeneinanders von Facetten, die die *Persönlichkeit* abbilden, dargestellt wird (Haller/Müller 2006), allerdings haben sich derartige Abgrenzungsversuche in der gesichteten Literatur nicht durchgesetzt.

objektivierenden und gestaltenden Subjekt aus, dessen Grundlage ein stabiles und sicheres Persönlichkeitsfundament ist. Dieses benötigt ihm zufolge jedoch Schutz gegen Angriffe von außen, um sich seinem Grundplan folgend zu einem funktionierenden mit sich selbst identischen Ganzen hin zu entwickeln. Durch Reflexion erfährt die Person dabei Gleichheit mit sich und diese als auch von anderen anerkannt (vgl. Paetow 2004, Haller/Müller 2006). Mead (1973) baute auf dieser Theorie Eriksons sein Konzept des Menschen auf, das auf der Idee der Unterscheidung zwischen einem empirischen *Me* und dem empirisch nicht erfassbaren *I* beruht. Während innerhalb dieses theoretischen Menschenbildes das *Me* in Form internalisierter Vorstellungen und Erwartungen des eigenen Fremdbildes auftritt, wobei alle *Me*'s zusammengenommen das *soziale Selbst* resp. die *soziale Identität*⁴⁹ bilden, gestaltet sich das *I* als das *reine Selbst*, das sich in seiner Kreativität und Spontanität äußert (vgl. Haller/Müller 2006). Dieser Dualismus hielt allerdings der Idee der wechselseitigen Wirkungen zwischen Subjekt und Umwelt nicht stand.

Um die konzeptionellen Überlegungen zum Persönlichkeitsbegriff abzurunden, hilft die Betrachtung des Persönlichkeitskonzepts Meyer-Drawes (2000): Diesem zufolge bleibt jede Identität (resp. Persönlichkeit) stets dynamisch, ohne sich gänzlich zu verfestigen, da sie sich in Rollen, Spiegelungen und Kontexten erst konstituiert. Folglich ist der eigene Ausdruck gegenüber anderen weder kontrollierbar, da dieser auf einer wechselseitigen Wahrnehmung und Interpretation beruht, noch manifest und daher sind die Möglichkeiten der eigenen Selbstreflexion begrenzt. Je nach Auffassung kann *Persönlichkeit* also als erlebte Einheit oder als multiples Konstrukt angesehen werden. Ungeachtet dessen muss in Bezug auf *Persönlichkeit*, sofern diese als das gilt, als was ein Mensch sich in der Welt zeigt, von einer *inneren* und einer *äußeren Persönlichkeit* ausgegangen werden. In dieser Arbeit wird in Anlehnung an Meyer-Drawe (2000) von einem Persönlichkeitsverständnis ausgegangen, das zwar auf dem Verhalten des Subjekts in der Welt basiert, dabei jedoch nicht von einer Identität von Erleben (im Sinne von Fühlen/Empfinden und Denken) und Handeln ausgeht und zudem zwischen subjektiver Selbstwahrnehmung und subjektiven Fremdwahrnehmungen

⁴⁹ Nach Goffman (1975) definiert sich die soziale Identität durch verschiedene Rollenidentitäten des Subjekts, während sich die davon abzugrenzende persönliche oder personale Identität durch individuelle Faktoren (wie Name, Aussehen, Eigenschaften) und die Biographie auszeichnet (Haller/Müller 2006).

differenziert. Diese Feststellung erfordert Respekt vor dem *Inneren* der Persönlichkeit ohne Einmischung: Die Persönlichkeit zeigt sich zwar in allen Äußerungen und sozialen Handlungen, deren Bedeutung geht jedoch immer sowohl über die Absichten als auch das Sichtbare hinaus (Schäfer 2005).

3.1 Forschungslinien und Forschungstraditionen der Persönlichkeitsforschung

Paradigmen sind verschiedene Forschungslinien, die sich gemäß ihrer Menschen- und Weltbilder in Theorien, Leitsätzen, Fragestellungen und Methoden mehr oder weniger voneinander unterscheiden und als Alleinstellungsmerkmale einander weitgehend ausschließen. Zu den Paradigmen der Persönlichkeitspsychologie zählen das dispositionale resp. das Eigenschaftsparadigma, das kognitive resp. konstruktivistisch-informationsverarbeitende Paradigma und das dynamisch-interaktionistische Paradigma, dem sowohl Behaviorismus als auch Determinismus zuzuordnen sind; ebenfalls wird das Konstrukt der *Persönlichkeit* unter biologischen Gesichtspunkten wie dem neurowissenschaftlichen, molekulargenetischen oder evolutionspsychologischen Paradigma sowie unter dem motivationspsychologischen, lerntheoretischen, humanistisch-bildungstheoretischen und psychoanalytischen Paradigma untersucht (vgl. Pervin 1987, Asendorpf/Neyer 2012, Asendorpf 2014, Rauthmann 2017); zudem – mitunter mit anderen Begriffsverwendungen – vor dem Hintergrund sozialpsychologischer und soziologischer Theorien (Haller/Müller 2006). Im Vergleich fallen die paradigmatischen Zuordnungssysteme mit ihren impliziten Strömungen insgesamt je nach Autor eher uneinheitlich aus (vgl. Pervin 1987, Haller/Müller 2006, Asendorpf/Neyer 2012, Asendorpf 2014, Rauthmann 2017).

Grundsätzlich lassen sich dennoch wenige Verknüpfungen unter den verschiedenen Paradigmen feststellen (Haller/Müller 2006), was dem Wesen eines Paradigmas innewohnt. Darüber hinaus galt es sogar neben teils grundlegend verschiedenen Paradigmen der Persönlichkeitsforschung zweitweise zwei Traditionen innerhalb der empirischen Psychologie zu unterscheiden: die Persönlichkeitspsychologie und die Differentielle Psychologie. Während im Fokus der Persönlichkeitspsychologie die Einzigartigkeit der (sich aus einzelnen Dimensionen ergebenden) Struktur menschlicher Persönlichkeiten stand, konzentrierte sich die Differentielle Psychologie auf die (sich in einzelnen Dimensionen manifestierenden) interindividuellen Unterschiede im Erleben und Verhalten. Allerdings ließ

sich diese Differenzierung nicht aufrechterhalten, da sich erst im Vergleich mit anderen die Einzigartigkeit des Einzelnen zu offenbaren vermag, weshalb mittlerweile beide Traditionen gängigerweise in einem Doppelnamen zusammengefasst werden. Da dieser psychologische Ansatz die Operationalisierbarkeit von Persönlichkeit anhand erfassbarer Merkmale voraussetzt, wird er mit diesem Fokus innerhalb des eigenschaftstheoretischen Paradigmas verortet (vgl. Asendorpf 2014, Rauthmann 2017). Es wird hierbei deutlich, dass es wohl weniger um tatsächliche, gefestigte Paradigmen als vielmehr um verschiedene Strömungen innerhalb von theoretischen Ansätzen geht, die vornehmlich auf ihnen entsprechenden Subjektvorstellungen basieren. Innerhalb eigenschaftstheoretischer Konzepte steht die Abbildung persönlicher Wesenszüge im Sinne „eine[r] starke[n] Neigung, sich in einer bestimmten Art und Weise zu verhalten“ (Pervin 1987, S. 287) im Vordergrund.

3.2 Operationalisierung des Konstrukts *Persönlichkeit*

Die Variation untersuchbarer persönlicher Wesenszüge resp. Persönlichkeitsmerkmale, die Individuen voneinander unterscheiden⁵⁰, ist stark ausgeprägt (vgl. Pervin 1987, Weber/Rammsayer 2012, Specht/Egloff/Schmukle 2011). Die Merkmale, die den Gegenstand der jeweiligen Persönlichkeitsforschung bilden, leiten sich meistens aus den (großen) Persönlichkeitstheorien⁵¹ ab und lassen sich nach Mischel (2004) innerhalb eines Kategoriensystems (dem sogenannten kognitiv-affektiven Persönlichkeitssystem) in fünf Kategorien systematisieren: 1. Enkodierung, 2. Erwartungen und Überzeugungen, 3. Affekte, 4. Ziele und Werte, 5. Kompetenzen und Fähigkeiten zur Selbstregulation (Weber/Rammsayer 2012).⁵² Noch einmal anders und weiter übergeordnet lassen sich

⁵⁰ U.a. Alexithymie, Bindungsstil, Cholerik, Depressivität, Empathie, Extraversion, Feindseligkeit, Gewissenhaftigkeit, Hochbegabung, internale Kontrollüberzeugung, Ja-Sage-Tendenz, kognitive Vermeidung, Lebenszufriedenheit, Machtmotiv, Neurotizismus, Offenheit für neue Erfahrungen, Optimismus, Psychotizismus, Regulativer Fokus, Selbstaufmerksamkeit, Trait Emotional Intelligence, Unterdrückung von Gefühlen, Verträglichkeit, Willensstärke (Weber/Rammsayer 2012).

⁵¹ Diese werden in der einschlägigen Literatur respektive als Subjekt- oder Identitätstheorien bezeichnet.

⁵² Zur Ergänzung: „Enkodierung - konstruktdifferente Strukturierung von sich selbst und seiner Umwelt; Erwartungen und Überzeugungen - interindividuelle Variation in Bezug

die Themenbereiche der Forschung zur Persönlichkeit folgendermaßen einordnen: a) in Kompetenzen, b) in emotional-motivationale Eigenschaften, c) in das Sozialverhalten sowie d) deren Korrelationen untereinander (Asendorpf 2014).⁵³ Die Methoden zur Erfassung des theoretischen Konstrukts *Persönlichkeit* variieren ebenso breit gefächert⁵⁴ (Asendorpf 2014). Diese Variation an Erhebungen zur Erfassung von Persönlichkeit erfolgt via Gegenstand und Inhalt, Datenquelle, Erhebungskontext, zeitliche Perspektive, Realitätsstatus sowie Antwortgestaltung (Weber/Rammsayer 2012). Darüber hinaus lassen sich sogenannten Wesenszüge oder Eigenschaften hierarchisch auf verschiedenen Niveaus verorten: Dabei gilt, dass einzelne Elemente der Persönlichkeit und des damit einhergehenden Verhaltens strukturell miteinander verbunden sind, wobei einige Elemente über einen stärkeren Einfluss verfügen als andere und damit „spezifische Reaktionen“ (Pervin 1987, S. 290) hervorrufen. Jene spezifischen Reaktionen, die im Sinne einer Eigenschaft allgemein miteinander verbunden sind – wie etwa der Gefallen an einer lebhaften Party mit vielen Menschen, was ein gewisses Maß eines geselligen Wesenszuges voraussetzt – lassen sich in Verbindung anderer Eigenschaften wiederum zu Dimensionen höheren Niveaus zusammenfassen, die ihrerseits wiederum in eine Typenbildung münden können. Problematisch am Niveau der Typenbildung ist allerdings zum einen, dass es jenseits des tatsächlich Beobachtbaren liegt und demnach stark vom spezifischen Verhalten abstrahiert wird. Zum anderen bedarf es zur Typenbildung stabiler Zusammenhänge vieler Reaktionen, die sich empirisch allerdings kaum bestätigen lassen (vgl. Pervin 1987). Ein stets offener Aspekt in der Debatte um die Erfassung von Persönlichkeitseigenschaften bleibt die Situationsspezifität von Verhalten und damit die Frage nach der Konsistenz von Verhalten. Dass sich diesbezüglich keine allgemeingül-

auf Zusammenhänge; Affekte - Gefühlsunterschiede; Ziele und Werte - u.a. Motive; Kompetenzen und Fähigkeiten zur Selbstregulation - Verhaltensrepertoire, Intelligenz“ (vgl. Weber/Rammsayer 2012, S. 13f).

⁵³ Kompetenzen (als leistungsermöglichende Eigenschaftsdeterminanten), emotional-motivationale Eigenschaften (dem Erleben und Verhalten zugrundeliegende Determinanten), das Sozialverhalten (in Interaktionen wirksam werdende Eigenschaftsdeterminanten).

⁵⁴ Die Variation rangiert zwischen idiographischen und nomothetischen Ansätzen, zwischen Selbstberichten, Verhaltensbeobachtungen und objektiven psychologisch-diagnostischen Verfahren, experimentellen Methoden, biologischen Ansätzen wie der Verhaltensgenetik, Zwillings- und Adoptionsstudien, Quer- und Längsschnittstudien sowie kulturvergleichenden Ansätzen (Asendorpf 2014).

tige Antwort finden lässt, da sowohl verschiedene Personen generell unterschiedlich konsistent in ihren Wesenszügen sind als auch bestimmte Situationen eher Konsistenz fördern als andere, muss gleichzeitig von einer Stabilität und einer situationsspezifischen Variabilität von Persönlichkeitseigenschaften ausgegangen werden (vgl. Pervin 1987, Rauthmann 2017).

Erste Versuche einer sogenannten objektiven Erfassung von Persönlichkeit lassen sich bereits zu der Zeit um 319 Jahre v. Chr. von Aristoteles' Schüler Theophrastos von Eresos verzeichnen (McCrae/Weiss 2007). Er fasste dabei beobachtbare Charakterzüge im Sinne von Verhaltensregelmäßigkeiten zusammen, ohne sich auf dahinter liegende Motive zu beziehen. Lange Zeit erfolgten derartige Beschreibungen von Persönlichkeit ausschließlich anhand von (mehr oder minder professionellen) Beobachtungen – was der eingangs beschriebenen Herleitung des Persönlichkeitsbegriffs entspricht. Erst sehr viel später, im 20. Jahrhundert – u.a. mit dem Wandel von Subjektvorstellungen und der Entwicklung der Idee eines von außen nicht einsehbaren Inneren des Menschen –, wandelte sich das Vorgehen vermehrt von der Fremdeinschätzung zur Selbsteinschätzung (McCrae/Weiss 2007). Mit dem während der Aufklärung aufkeimenden Gedanken der Autonomie des Subjekts, genährt von psychoanalytischen Vorstellungen des Subjekts als *Herr im eigenen Hause* (vgl. Meyer-Drawe 2000), werden heute trotz der historischen Bedeutsamkeit von (Beobachter- resp.) Fremdeinschätzungen vornehmlich Selbsteinschätzungen zur Persönlichkeitsbeurteilung eingesetzt (McCrae/Weiss 2007, Furr 2009, Baumeister/Vohs/Funder 2007). Zudem erfolgen die Qualitäts- und Validitätsbeurteilung von Fremdeinschätzungen in der Regel entlang ihrer Übereinstimmung mit der Selbsteinschätzung (McCrae/Weiss 2007), d.h. die Selbsteinschätzung der Persönlichkeit gilt heute vorrangig als Maßstab.⁵⁵

Dieser Umstand offenbart sich auch in nahezu allen Untersuchungen, die das Konstrukt *Persönlichkeit* thematisch implizieren: es erscheint oft nicht einmal erwähnenswert, aus welcher Perspektive *die Persönlichkeit* erfasst wird. Ange-

⁵⁵ Eine Ausnahme bildet dabei allerdings die Beurteilung des Temperaments und der Persönlichkeit von Kindern, zu der grundsätzlich und fast ausschließlich Beobachtereinschätzungen herangezogen werden (McCrae/Weiss 2007). Ebenso lässt sich in Bezug auf psychische Erkrankungen und deren Diagnose primär der Einsatz professioneller Fremdbeurteiler feststellen (Horsch 2015). Die Möglichkeit der validen Persönlichkeitsselbsteinschätzung erscheint demnach an die (per Fremdbeurteilung) potentielle oder zugesprochene Mündigkeit eines Menschen gebunden.

sichts der dieser Arbeit zugrundeliegenden besonderen Bedeutung der Perspektivität bedarf es jedoch Möglichkeiten der empirischen Erfassung von Persönlichkeit, die sowohl Fremd- als auch Selbsteinschätzungen zulassen. Das bedeutet, dass – um der Fremdeinschätzung gerecht zu werden – eigenschaftszentrierte Persönlichkeitstheorien und deren adäquate Operationalisierung relevant werden. Zur Erfassung von Persönlichkeitseinschätzungen entlang von Merkmalen eignen sich vornehmlich standardisierte und validierte Instrumente (McCrae/Weiss 2007). Da in dieser Hinsicht der BigFive-Ansatz das derzeit gängigste und etablierte Konzept zur Erfassung der Persönlichkeit darstellt (Haller/Müller 2006), dessen Modell sich sowohl in der Selbst- als auch der Fremd-einschätzung bewährt hat (Asendorpf/Neyer 2012), kulturübergreifend weitgehend bestätigt wurde, alltagspsychologische Akzeptanz genießt und bei Kindern, Jugendlichen und Erwachsenen sowie situationsspezifisch in unterschiedlichsten Kontexten funktioniert, erscheint es angemessen, sich insbesondere auf dieses Konzept zu konzentrieren. Das Modell der BigFive erfragt zudem derart umfassend die Persönlichkeitseigenschaften, dass sich zu den meisten relevanten alternativen Anwendungen entsprechende Skalen darin widerspiegeln (McCrae/Weiss 2007). Darüber hinaus hat das Konstrukt der BigFive den Vorteil, dass es nicht auf Typenbildung basiert (wie AVEM (Arbeitsbezogene Verhaltens- und Erlebensmuster) oder MBTI (Meyers-Briggs Type Indicator)), wodurch eine höhere Differenziertheit der einzelnen erfassten Merkmale erhalten bleibt (vgl. u.a. Furnham 1996).

3.3 BigFive der Persönlichkeit

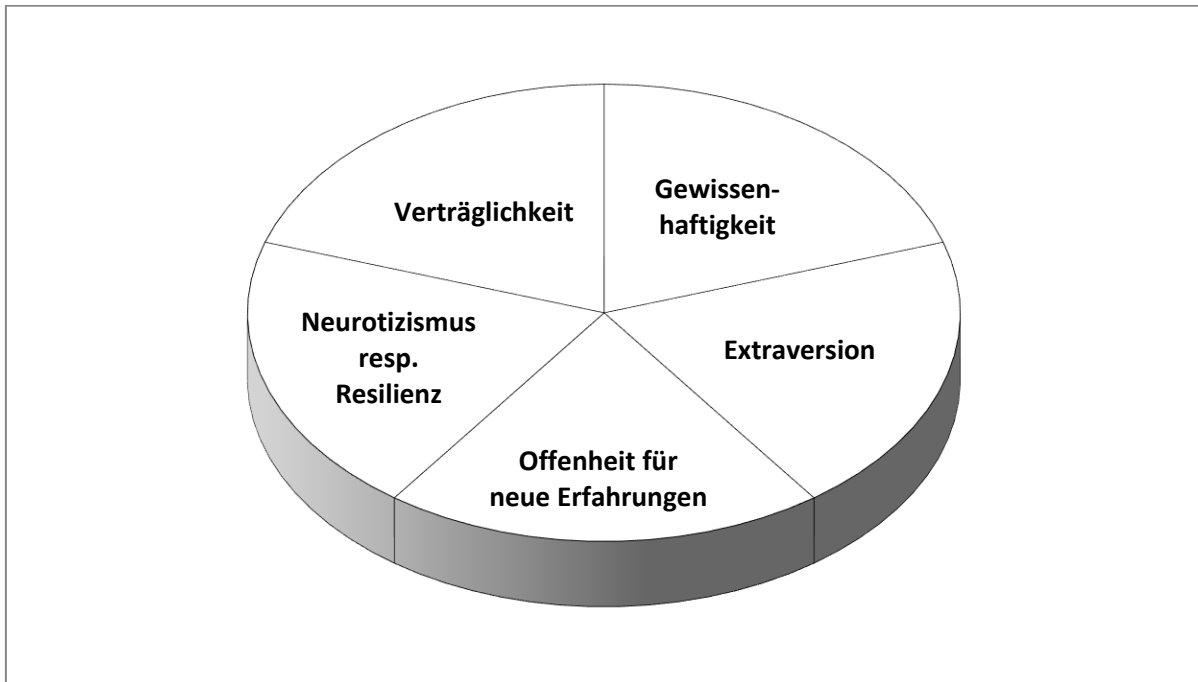


Abb. 2: Modell nach Costa/McCrae 1989.

Der BigFive-Ansatz gründet auf der Annahme, „dass sich alle Aspekte der menschlichen Persönlichkeit, die in alltäglichen Verhaltensweisen zwischen Menschen von Bedeutung sind, im Vokabular einer Sprache wiederfinden.“ (Haller/Müller 2006, S.14, De Raad 2000, S. 16). Diese Annahme entspricht der lexikalischen Hypothese de Raads (vgl. De Raad 2000 in Haller/Müller 2006). Davon ausgeschlossen sind auf Verhaltensdispositionen anspielende Begriffe, die bewertende Bezeichnungen, Einstellungen, Werthaltungen oder körperliche Merkmale benennen (Asendorpf/Neyer 2012). Gleichzeitig wird das Individuum als zeitlich relativ stabil durch die spezifische Kombination der verschiedenen Merkmale gekennzeichnet aufgefasst (Haller/Müller 2006), ohne jedoch eine Klassifikation von Typen⁵⁶ vorzunehmen. Zwar sind Typen-Klassifikationen für Laien und Anwender fremder Fachgebiete in der Regel leichter zu verorten, gegenüber graduell abgestuften Beschreibungen anhand von Profilen allerdings derart vereinfachend, dass sie Gefahr laufen, zu einem auf Werten basierenden Schubladendenken zu verkommen (Asendorpf/Neyer 2012). Demnach be-

⁵⁶ Typen stehen im Gegensatz zu Merkmalen auf einer höheren Organisationsebene, auf der die Merkmale zu spezifischen Typen verbunden werden (Pervin 2000).

schränkt sich das Konzept der BigFive auf eine Klassifikation von Eigenschaften, nicht aber von Personen (Hossiep 2014). Diese nach den oben genannten Kriterien und auf einem hohen Abstraktionsniveau durch Beobachtung oder Reflexion erfassbaren Eigenschaften münden in diesem Konzept in fünf übergeordnete Merkmalsdimensionen *Gewissenhaftigkeit*, *Extraversion*, *Offenheit für neue Erfahrungen*, *Neurotizismus* und *Verträglichkeit* – die BigFive (vgl. Backhaus 2004).

Kritik an BigFive-Instrumentarien basiert auf den hohen Skaleninterkorrelationen und der oftmals schwach ausgeprägten Validität (z.B. auch bei Ostendorf/Angleitner 2004, Moser/Soucek 2003), die einerseits der breit gefassten Definition von Persönlichkeitseigenschaften (vs. eng gefassten Persönlichkeitsvariablen) und andererseits spezifischen Beurteilungstendenzen geschuldet sind (Hülshager/Splecht/Spinath 2006, Asendorpf/Neyer 2012). Eine Verbesserung der Validität kann durch kontextbezogene Itemformulierungen erfolgen, wodurch ein gemeinsamer Bezugsrahmen der Testpersonen geschaffen wird, was zur Stärkung des Zusammenhangs von BigFive und bspw. der Arbeitsleistung beim Einsatz des Bezugsrahmens *Arbeit* (im Gegensatz zu einer allgemeinen Erhebung) führt (Hunthausen/Truxillo/Bauer/Hammer 2003).⁵⁷

3.3.1 Entwicklung der BigFive

Die BigFive sind fünf breit angelegte, zeitlich und kulturell überdauernde Persönlichkeitsmerkmale bzw. Merkmalsdimensionen, die zugleich auf eine lexikalische und eine empirische Entwicklung zurückzuführen sind (Hossiep 2014). Diese beiden Entwicklungslinien sind derart miteinander verschränkt, dass ihre Entwicklungsgeschichten innerhalb der Literatur mehr oder weniger weit auseinandergehen. Die lexikalische Entwicklung nach dem psycholexikalischen Ansatz stützt sich auf Ähnlichkeiten in der Verwendung von Eigenschaftsbegriffen im Alltag, d.h. auf persönlichkeitsbezogenes Vokabular im Sinne überdauernder Eigenschaften, die begrifflich durch Adjektive erfassbar sind (Backhaus 2004, Häcker 2014). Dabei wird Persönlichkeit als eine Gesamtheit aller nicht-situativen Verhaltensbedingungen, die Verhalten prognostizieren, verstanden (Häcker

⁵⁷ Entlang dieses Prinzips wurde bspw. der BIP (Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung) entwickelt (Hossiep/Paschen 2003, Hülshager/Splecht/Spinath 2006).

2014). Konkret wurden also alle personenbeschreibenden Adjektive des Lexikons, die weder wertend sind noch auf Einstellungen, Werthaltungen oder körperliche Merkmale abzielen, einzeln herausgesucht. Diese Arbeit erfolgte 1936 durch Allport und Odberg, die nach der Durchsicht von Wörterbüchern eine Liste mit 17.953 Begriffen erstellten und die Fachwelt zu einer logisch stringenten Reduktion dieser Liste aufriefen (vgl. Allport 1937, Haller/Müller 2006). Dieser Aufgabe widmete sich Cattell und generierte 1946 eine auf 16 grundlegende Eigenschaften reduzierte Liste, den sogenannten Primärfaktoren (Backhaus 2004). Nahezu zeitgleich identifizierte 1947 Eysenck innerhalb der klinischen Psychologie anhand von Fragebogenanalysen auf der Basis von beschreibenden Aussagen die Eigenschaften *Neurotizismus* und *Extraversion* als zentrale Merkmale der Persönlichkeit (Backhaus 2004). Ausgangspunkt seiner Analysen ist die theoretische Annahme, dass einzeln beschreibbare Verhaltensweisen resp. Wesenszüge systematisch zusammengefasst ein Verhalten im Sinne eines Typenniveaus beschreiben. Demzufolge bilden spezifische Reaktionen insgesamt habituelle Reaktionen ab, aus denen sich Merkmalen ergeben, die wiederum auf einer höheren Organisationsebene zu Persönlichkeitstypen verbunden werden können (Pervin 2000, Häcker 2014).

Anhand des lexikalischen Ansatzes wurde Cattells Liste an Eigenschaftsbegriffen durch Norman weiter reduziert, bis 1967 eine auf 2.800 Begriffe reduzierte Liste entstand (Haller/Müller 2006). Entlang der Idee von Primärfaktoren von Cattell reduzierten Costa und McCrae 1980 diese Liste mittels Faktorenanalyse und fanden dabei den Befunden Eysencks entsprechend die Eigenschaften *Neurotizismus* und *Extraversion* als Primärfaktoren, die sie Dimensionen nennen, sowie die weiteren Dimensionen *Offenheit für Erfahrungen* (*openness to experience*), *Verträglichkeit* (*agreeableness*) und *Gewissenhaftigkeit* (*conscientiousness*) (Backhaus 2004). Im Weiteren fasste Goldberg 1990 erneut 339 Eigenschaftswörter zu 100 Gruppen zusammen, die nach Durchführung verschiedener Tests zur Selbst- und Fremdbeurteilung einer Faktorenanalyse unterzogen wurden. Die Faktorenanalyse offenbarte ebenfalls die Extraktion von fünf Faktoren (Haller/Müller 2006). Demnach gelten heute die BigFive als psycholexikalische Entwicklung, die anhand des Fragebogenansatzes der differentiellen-klinischen Psychologie modifiziert wurde (durch Einbeziehung der Eigenschaften *Neurotizismus* und *Extraversion* als eigenständige Dimensionen), sie sind also sowohl qualitativ als auch quantitativ entwickelt (Backhaus 2004, Lang/Lüdtke 2005).

Innerhalb des Modells der BigFive mit den fünf großen, übergeordneten Dimensionen lassen sich Subgruppen ausmachen, die ihrerseits gewisse Facetten der Dimensionen widerspiegeln. Jede der fünf Dimensionen beruht demnach wiederum auf sechs solcher Facetten. Die personenbeschreibenden Adjektive wurden also innerhalb von dreißig Facetten eingeordnet, wobei sechs Facetten jeweils eine Dimension abbilden, die ihrerseits somit je einen Faktor zweiter Ordnung darstellen (vgl. Backhaus 2004). Im folgenden Kapitel wird diese hierarchische Struktur detailliert beschrieben.

Dennoch ist die Anzahl an *tatsächlichen*, offenbar relevanten Faktoren, die notwendig sind, um Persönlichkeit möglichst präzise abzubilden, ebenso strittig wie die den Faktoren zugeordneten Begriffe und deren inhaltliche Interpretation (Hossiep 2014, Haller/Müller 2006). So stellt es ein gängiges Dilemma in der Persönlichkeitskonstruktion dar, ob zur Persönlichkeitserfassung drei, vier, sechs oder auch mehr Dimensionen resp. Faktoren notwendig sind (Haller/Müller 2006, Asendorpf/Neyer 2012), obgleich die fünffaktorielle Struktur in einer Vielzahl an Studien belegt wurde (Backhaus 2004). So zeigt sich beispielsweise, dass es sieben Faktoren (*BigSeven*) bedarf, sofern – entgegen des dem BigFive-Konzept zugrundeliegenden Verständnisses (vgl. S. 46f) – wertende Bezeichnungen, Einstellungen und Werthaltungen unter den analysierten Begriffen einbezogen werden. Und obgleich die Fünffaktorenstruktur auch als (bzgl. des englischen, deutschen und niederländischen Raums (siehe Asendorpf/Neyer 2012)) kulturübergreifend bestätigt gilt (Hossiep 2014), zeigt sich bei genauerer Betrachtung, dass vor allem drei große Dimensionen (*BigThree*) – *Extraversion*, *Gewissenhaftigkeit*, *Verträglichkeit* – kulturübergreifend stabil abbildbar sind (De Raad 2010, Asendorpf/Neyer 2012). Bezüglich der faktorenanalytischen Zusammenfassung von Begriffen bleibt vor allem offen, wie sich der fünfte Faktor tatsächlich zusammensetzt, da sich dieser über verschiedene Länder hinweg durchgehend heterogen darstellt: Teils wird der Faktor anders benannt und definiert oder aber er wird mit anderen Facetten belegt (De Raad 1998, Backhaus 2004).⁵⁸

Demnach gilt – aus Sicht idiographischer Orientierung – die auf Faktorenanalysen beruhende Eigenschaftstheorie als Paradigma angreifbar, da sie sich nicht abschließend generalisieren lässt (Häcker 2014). Die Ausprägungen der Faktoren

⁵⁸ Die getesteten Sprachen sind Englisch, Deutsch, Dänisch, Ungarisch, Italienisch, Tschechisch, Polnisch, Chinesisch, Portugiesisch, Quechua, Russisch, Rumänisch, Spanisch, Philippinisch, Arabisch und weitere (De Raad 1998).

beruhen möglicherweise zu sehr auf alltagspsychologischen Wahrnehmungen, die wissenschaftlichen Klassifikationen nicht adäquat standhalten. Zudem können generalisierende Aussagen über Verhaltensweisen dadurch erschwert sein, dass das Verhalten zwischen Settings und Situationen variiert. So wird bspw. Ärger im Privaten möglicherweise direkt ausgedrückt, was in einem beruflichen Kontext unterbleibt (Weber/Rammsayer 2012, mehr dazu in Kap. 3.4.2). Trotz derartiger Kritik gilt das innerhalb eines fünfdimensionalen Koordinatensystems variablenorientierte Beschreibungssystem der BigFive, das anhand von Profilen graduell abgestufte Persönlichkeitsbeschreibungen ermöglicht, als das derzeit beste existente Modell zur Erhebung der Persönlichkeit (De Raad 1998).

3.3.2 Systematik des BigFive-Konstrukts: Dimensionen/Facetten

Die Dimensionen der BigFive sind jeweils nach einem Pol eines dualen Kontinuums benannt, implizieren jedoch immer auch den Gegenpol (siehe Tab. 2). Besonders bedeutsam ist dabei, dass die jeweilige Ausprägung prinzipiell wertfrei zu interpretieren ist, wenngleich sie nicht vor Verzerrungen wie der sozialen Erwünschtheit gefeit ist.

Tab. 2: *BigFive-Dimensionen mit ihren Facetten (adaptiert aus Ostendorf/Angleitner 2003, ergänzt anhand von Simon 2006).*

Dimension	Gewissenhaftigkeit	Extraversion	Offenheit für Erfahrungen	Neurotizismus	Verträglichkeit
Facetten	Kompetenz	Herzlichkeit	Phantasie	Ängstlichkeit	Vertrauen
	Ordnungsliebe	Geselligkeit	Ästhetik	Reizbarkeit	Freimütigkeit
	Pflichtbewusstsein	Durchsetzungsfähigkeit	Gefühle	Depression	Altruismus
	Leistungsstreben	Aktivität	Handlungen	Soziale Befangenheit	Entgegenkommen
	Selbstdisziplin	Erlebnishunger	Ideen	Impulsivität	Bescheidenheit
	Besonnenheit	Frohsinn	Werte	Verletzlichkeit	Gutherzigkeit
Gegenpol	Nachlässigkeit	Introversion	Konservatismus, Unbeweglichkeit	Resilienz	Konkurrenz, Antagonismus

Die folgenden Beschreibungen der BigFive-Dimensionen sind an Simon (2006) und Asendorpf/Neyer (2012) angelehnt (die expliziten Attribuierungen der einzelnen Facetten werden in der Vorstellung des BigFive-Erhebungsinstruments in Kap. 11.2 beschrieben):

Gewissenhaftigkeit impliziert die Kontrolliertheit im Sinne der Fähigkeit, planvoll und organisiert zu handeln, was jedoch bis hin zu *verkopftem* Verhalten und zwanghafter Ordnungsliebe ausarten kann. Die Kontrolliertheit umschließt zudem mehrere Facetten der Gewissenhaftigkeit. Besonnenheit und Selbstdisziplin demgegenüber sind eher den Temperamenteigenschaften zuzuordnen. Insgesamt sind hohe Werte in der Dimension Gewissenhaftigkeit sozial erwünscht, extrem hohe Werte allerdings können auf sogenannte *kleinkarierte Pedanten* oder *Wor-*

kaholics hinweisen. Voraussetzung für hohe Gewissenhaftigkeit und kompetentes Handeln ist vor allem die Selbstdisziplin, was mit dem Befund harmoniert, dass Gewissenhaftigkeit der BigFive-Faktor ist, der Leistungen am besten vorhersagt. Entsprechend sagt die selbstbeurteilte Gewissenhaftigkeit auch das Vorgesetztenurteil über den Berufserfolg am besten vorher.

Extraversion beschreibt nur teilweise Temperamentsmerkmale wie Aktivität, Erlebnishunger und Frohsinn. Interpersonelle Stile im Sinne stabiler Verhaltens-tendenzen in sozialen Interaktionen äußern sich bei hohen Werten in dieser Dimension in Form von Herzlichkeit, Geselligkeit, Freude an Kommunikation und Durchsetzungsfähigkeit. Niedrige Werte weisen auf Introversion als das Fehlen extravertierter Persönlichkeitsanteile hin. Da extravertierte Eigenschaften leicht von außen beobachtbar sind und sich vor allem in den interpersonalen Stilen zeigen, erklärt dies möglicherweise die hohen Übereinstimmungen (Korrelationen um 0,5) zwischen Selbst- und Fremdbeurteilung (mehr dazu in Kap. 3.4.1) innerhalb dieser Dimension.

Offenheit für neue Erfahrungen in verschiedenen Bereichen (siehe Facetten) steht vor allem für intellektuelle Neugier und ästhetisches, künstlerisches Empfinden. Auch Kreativität und vielseitige, gelebte Emotionalität werden innerhalb dieser eher abstrakt und liberal ausgerichteten Dimension verortet. In diesem Sinne äußert sich diese Dimension in einer Offenheit gegenüber Neuerungen, Abwechslungen und Vielfalt sowie die Infragestellung bestehender Werte und Normen. Hohe Werte zeigen sich gehäuft in Verbindung mit hoher Intelligenz und einem hohen Bildungsgrad.

Neurotizismus als Empfindlichkeit neurotischer Menschen wird oft als *interessant-schillernd* wahrgenommen, wenn sie mit Empfindsamkeit einhergeht, dennoch sind hohe Werte von Neurotizismus sozial eher unerwünscht. Während niedrige Werte zwar sozial erwünscht sind, wird niedriger Neurotizismus oftmals als Insensibilität wahrgenommen, was sich dadurch erklären lässt, dass belastbare Personen eher stärkerer und mehr Reize bedürfen, um Emotionen wahrzunehmen, während sensible Personen empfindlich auf Reize reagieren. Neurotizismus kann insgesamt kaum von Ängstlichkeit unterschieden werden (und zeigt diesbezüglich sehr hohe Korrelationen), entsprechend gilt soziale Ängstlichkeit auch als neurotische Introversion. In der Arbeitswelt sagt die Dimension Neurotizismus vielfältige Probleme vorher, wie beispielsweise mangelnden Einsatz durch schnell einsetzende subjektive Überforderung, Probleme im Kundenkontakt oder in Teams – die Probleme zeigen sich vor allem durch hohe Reizbarkeit

und Verletzlichkeit sowie häufige Wechsel des Arbeitsplatzes. Diese Dimension wird in der Regel anhand ihres mitunter eher als negativ konnotierten Pols *Neurotizismus* betitelt, wohingegen die anderen vier Dimensionen anhand ihrer eher positiven Ausprägungen betitelt werden. Im weiteren Verlauf der Arbeit wird die *Neurotizismus*-Dimension daher meist als *Resilienz* bezeichnet, um diese Dimension äquivalent zu den anderen Dimensionen zu behandeln.

Verträglichkeit bezieht sich vornehmlich auf soziale Motive, nicht auf Temperamenteigenschaften. Diese Dimension wird demnach mit Freundlichkeit und Hilfsbereitschaft verbunden, die auf Wärme im zwischenmenschlichen Umgang basieren. Damit einhergehend sind hohe Werte größtenteils sozial erwünscht, jedoch wird übermäßiges Vertrauen und Entgegenkommen wiederum eher als naiv und unterwürfig gewertet. Niedrige Werte sind dennoch durchweg sozial unerwünscht und können im Extremfall sogar Persönlichkeitsstörungen anzeigen.

Zwar stellen sich die BigFive als separat zu betrachtende Dimensionen dar, jedoch sind die Ausprägungen der einzelnen Dimensionen gegebenenfalls im Gesamtprofil zu deuten: hohe neurotizistische Ausprägungen beeinträchtigen bspw. das Antwortverhalten bzgl. der anderen Dimensionen; hohe extravertierte Ausprägungen bspw. ergeben unterschiedliche Bilder, je nachdem, ob sie mit bspw. einer ebenso hohen Verträglichkeit einhergehen oder nicht (vgl. Simon 2006). Hinsichtlich interpersoneller Stile wiederum ist vor allem die Differenzierung innerhalb eines zweidimensionalen Modells, das durch die Dimensionen *Extraversion* und *Verträglichkeit* bestimmt ist, aussagekräftig (vgl. Asendorpf/Neyer 2012). Diese Befunde sprechen für eine essentielle Multikollinearität⁵⁹ zwischen den Dimensionen, wodurch Regressionseffekte mit Vorsicht zu interpretieren sind (mehr dazu in Kap. 16.2 und 18.3).

3.4 Forschungsüberblick zur Persönlichkeit

Die ursprüngliche englischsprachige Version eines Erhebungsinstruments – der NEO-FFI (Five Factor Inventory) – stammt von Costa und McCrae (1989) und wurde von Borkenau und Ostendorf (1993) für den deutschsprachigen Raum

⁵⁹ Eine essentielle Multikollinearität liegt dann vor, wenn verschiedene prädiktive Variablen in einer Regressionsanalyse „substantiell bzw. inhaltlich hoch miteinander korrelieren“ (Urban/Mayerl 2011, S. 235). Daher ist sie auch kaum zu bereinigen. Sie ist von einer nicht-essentiellen, d.h. im Forschungsdesign angelegten linearen Abhängigkeit, zu unterscheiden (vgl. Urban/Mayerl 2011).

adaptiert. Er misst die BigFive anhand von 60 Items, d.h. mit zwölf Items pro Dimension, demnach mit lediglich zwei Items pro Facette (Haller/Müller 2006). Insgesamt muss in Bezug auf die Persönlichkeitsforschung beachtet werden, dass diese nicht allgemeine Gesetzmäßigkeiten, sondern lediglich eine differenzierte Betrachtungsweise anvisiert – demnach werden hinsichtlich des Konstrukts *Persönlichkeit* weder Absolutheitsansprüche erhoben noch auf letztgültiger Vollständigkeit bestanden (Schaffhuser 2014).

Nach der Darstellung von Befunden bezüglich der Eigenschaften der einzelnen Perspektiven und des Verhältnisses von Fremd- resp. Beobachtereinschätzungen und Selbsteinschätzungen entlang des BigFive-Konstrukts (Kap. 3.4.1) wird ersichtlich, dass diese auf eine perspektivenabhängige Validität der Einschätzungen hinweisen. Demzufolge stützen die empirischen Untersuchungen theoretische Annahmen, die von unterschiedlichen *Selbsten* ausgehen (siehe hierzu Kap. 3).

Da Fremdeinschätzungen allerdings noch immer unüblich sind und deren Einsatz wenig erforscht ist (Schaffhuser 2014), beziehen sich die allgemeinen Forschungsbefunde (Kap. 3.4.2, 3.4.3) auf Selbsteinschätzungen der Persönlichkeit. Darin schlägt sich jedoch v.a. die Interpretation eigener Verhaltensweisen via des Moderators *Selbstbild* (resp. *Me* gemäß Meads Theorieansatz, siehe Kap. 3) im Antwortverhalten nieder. So beeinflusst die Interpretation der eigenen Verhaltensweisen das Selbstbild und das Selbstbild wiederum wirkt auf das Antwortverhalten im Persönlichkeitstest (Backhaus 2004).

3.4.1 Das Verhältnis von Fremd- und Selbsteinschätzung der Persönlichkeit

„How others see us contains important information with respect to what we are like.“ (Back/Vazire 2012, S.134)

Persönlichkeitsmerkmale beruhen auf der Beschreibung des Denkens, des Fühlens und/oder des Handelns von Menschen, durch deren zeitlich relativ überdauernden Muster sich Individuen voneinander unterscheiden (McCrae/John 1992); sie beschreiben das individuelle Denken und Empfinden anderen Personen und sich selbst gegenüber und wie ein Individuum typischerweise die soziale Umgebung wahrnimmt und auf diese reagiert (Back et al. 2011, Caspi/Roberts 2001). Obwohl Merkmale der Persönlichkeit in weiten Teilen ebenso von außen, d.h. von anderen, wie von innen, d.h. sich selbst, zugeschrieben werden können,

beruht die Persönlichkeitsforschung bis dato fast ausschließlich auf Selbstberichten. Neben der Selbstwahrnehmung ist allerdings auch die Fremdwahrnehmung bedeutsam, um spezifische Informationen darüber zu erhalten, wie die individuelle Persönlichkeit in spezifischen Umgebungen ausgedrückt, wahrgenommen und eingeschätzt wird. Um die Differenzierung zwischen verschiedenen Perspektiven auf Persönlichkeit zu verifizieren, wurden in einer Studie⁶⁰ von Schaffhuser, Allemann und Martin (2014) Persönlichkeitseinschätzungen aus drei Perspektiven erhoben:

1. Selbstbericht – bietet einen einmaligen und reflexiven Blick auf die eigene Identität anhand intrapsychischer Informationen (Hogan/Roberts 2004, Roberts/Wood 2006). Allerdings bestehen zum einen introspektive Begrenzungen, sogenannte ‚Blinde Flecken‘ im Selbstbild (siehe auch Kap. 1.1), wodurch bspw. automatisierte Verhaltensweisen im Selbstbericht kaum direkt erfassbar sind (Back/Vazire 2012). Zum andern besteht bei dieser Perspektive eine hohe Verzerrungsgefahr, bspw. durch Selbstverstärkung, -aufwertung und -überschätzung, sozial erwünschtes Antwortverhalten und/oder Vergleichseffekte (Back/Vazire 2012, Srivastava 2012, Paulhus/Vazire 2007).
2. Metabericht – Aussage darüber, wie ein Individuum glaubt, dass andere es wahrnehmen. Als solche ist diese Perspektive stark beeinflusst von der Selbstwahrnehmung (Kenny/DePaulo 1993) und fällt damit eher unter die Perspektive des Selbstberichts.
3. Fremdbbericht – ist eine „Methode der Persönlichkeits- und Verhaltensbeurteilung durch mehr oder weniger informierte Beurteiler [...]“ (Neyer in Wirtz 2014, S. 609) und zeigt in spezifischen sozialen Kontexten wahrgenommene Persönlichkeitsmerkmale auf (Back et al. 2011, Hogan/Roberts 2004, Roberts/Wood 2006). Beurteiler gelten hier als Experten bestimmter, begrenzter Situationen (Asendorpf/Neyer 2012) und lassen sich in verschiedene Gruppen unterteilen. Dazu gehören Familie, Freunde, Mitbewohner, Bekannte, Arbeitskollegen und Fremde (Rauthmann 2017). Jede dieser Perspektiven eröffnet dabei den Blick auf typische Verhaltensmus-

⁶⁰ Die berichtete Studie fokussierte im Detail die Zufriedenheit von Paaren in Beziehungen, demnach stammen die erfassten Fremdbberichte von gut bekannten Dritten.

ter, die sich in sozialer Interaktion anderen sichtbar zeigen, ohne dem jeweiligen Individuum unbedingt selbst bewusst zu sein, d. h. Fremdbenachrichte liefern einmalige und relevante Informationen und gelten zudem als objektiver resp. weniger verzerrt als Selbstberichte sowie als reliabel und valide (Hofstee 1994, Vazire 2010) trotz Beobachterbias‘ (Weller/Watson 2009). Die Akkuratesses (im Sinne einer differenzierten Erfassung) einer Fremdeinschätzung hängt von verschiedenen Aspekten ab:

- a) Beurteilungskompetenz – d.h. vom (Vor-)Wissen, der individuellen Fähigkeit sowie der Motivation zur Einschätzung
- b) generelle Beurteilbarkeit – d.h. vom Bekanntheitsgrad mit der zu beurteilenden Person und deren Verhaltenskonsistenz
- c) zu beurteilende Merkmale (und der damit einhergehenden Qualität der Information) – hierbei steht v.a. der Kontext der Beziehung im Vordergrund
- d) Beurteilungsfehler – diese gründen v.a. auf impliziten Persönlichkeitstheorien, die durch entsprechende, angemessene Itemkonstruktionen nur teilweise behebbar sind

Es können also in der Forschung grundsätzlich drei Erhebungsperspektiven unterschieden werden (vgl. Schaffhuser/Allemand/Martin 2014). Entlang dieser Perspektiven wurden verschiedene Modelle jeder einzelnen BigFive-Dimension erstellt, die jeweils unterschiedliche Anzahlen an Faktoren umfassten⁶¹. Dabei zeigte sich, dass ein Zwei-Faktorenmodell bereits besser die Daten erfasst als ein Single-Faktor-Modell, die besten Model-Fit-Werte jedoch das Drei-Faktoren-Modell aufweist (vgl. Kap. 10). Zwar zeigen die drei Perspektiven alle Interkorrelationen untereinander, stellen jedoch eindeutig getrennte Faktoren dar, d.h. ähnliche und dennoch eigene Perspektiven auf Persönlichkeit. Jede der Perspektiven beleuchtet dabei spezifische Aspekte der Persönlichkeit und unterliegt ihren eigenen Biasfaktoren⁶².

⁶¹ Alters- und Geschlechtseffekte wurden dabei als erstes in jedem Analyseschritt kontrolliert.

⁶² Sowohl bei der Selbst- als auch bei der Fremdbewertung bestehen gemeinsame Biaseffekte, z.B. der Halo-Effekt, anhand dessen eine Eigenschaft das Urteil über eine andere Eigenschaft beeinflusst und so möglicherweise zu Scheinkorrelationen führt. Allerdings gelten vor allem Faktoren 2. Ordnung als von Halo-Effekten verzerrt. Auch stellen sozial erwünschte Urteile einen Bias dar, da bei der Bewertung persönlicher Eigenschaften eher positive Eigenschaften hervorgehoben werden, was zu Selbst- und Fremdtäuschung führen kann (Asendorpf/Neyer 2012).

Durch die Aggregation mehrerer Fremdeinschätzungen ist es möglich, Urteilsverzerrungen zu reduzieren und die Reliabilität der erhobenen Merkmale im Sinne der internen Konsistenz (Cronbachs α , Intraklassenkorrelationen) zu erhöhen. Dafür sollten mindestens vier Urteile zur Aggregation vorliegen (nach Wirtz 2014). Fremdurteile, die auf Bekanntheit mit der einzuschätzenden Person aus derselben Situation basieren, ähneln einander zudem stärker als Urteile von Personen, die die einzuschätzende Person aus unterschiedlichen Kontexten kennen (Asendorpf/Neyer 2012). Demzufolge müssten bspw. Lehrerpersönlichkeitsurteile von Schülerinnen und Schülern innerhalb einer Klasse eher stark miteinander korrelieren. Darüber hinaus muss bei Fremdeinschätzungen zwischen den Perspektiven von Akteur und Beobachter unterschieden werden (Abele 2014). In sozialen Situationen sind Personen sowohl Akteure als auch Beobachter, d.h. ihre Wahrnehmung unterliegt einem ständigen Perspektivenwechsel mit Auswirkungen auf die Urteilsbildung. Nach Jones/Nisbett (1972) erfolgen diese Auswirkungen auf die Urteilsbildung durch Verzerrungen von Kausalattributionen, die ihrerseits wiederum auf Informations- und Wahrnehmungsunterschiede (sowie naive Psychologie) zurückzuführen sind. Zumal Malle (2006) zeigen konnte, dass eigene negative Verhaltensweisen eher mit dem Akteur-Beobachter-Unterschied erklärt werden als positive (vgl. hierzu auch ‚actor-observer-bias‘ in Wirtz 2014). Dadurch, dass unterschiedliche Parameter der Urteilsbildung je nach Perspektive eingesetzt werden, erfolgt die Beurteilung anderer über sogenannte *kommunale Eigenschaften*, die Aussagen über Intentionen erfassen; bei sich selbst hingegen erfolgt die Beurteilung über *agentische Eigenschaften*, die Aussagen über die konkrete Zielerreichung eines Aspekts beinhalten (Abele/Wojcizke 2007).

Überprüfungen zur Perspektivenspezifität von Eigenschaftsmessungen finden in der Regel über Korrelationen statt. Zwischen der Selbst- und der Metawahrnehmung zeigen sich dabei sehr starke Korrelationen (Gallrein/Carlson/Holstein/Leissing 2013). Dadurch wird die Annahme, dass Individuen denken, andere sähen sie so wie sie selbst sich sehen, als bestätigt angesehen (Kenny/DePaulo 1993). Wobei es zu beachten gilt, dass es als ein grundlegendes, starkes Bedürfnis gilt, von Anderen eine Bestätigung der Selbstwahrnehmung zu erhalten (Schaffhuser/Allemand/Martin 2014). Allerdings handelt es sich hierbei um den Vergleich zweier Selbstbilder und lässt demnach vor allem Rückschlüsse auf das Selbstkonzept zu (Schaffhuser/Allemand/Martin 2014). Über den Vergleich von Fremdbereichten mit Selbstberichten findet in der Regel eine Validitätsprüfung

statt, wobei die Kongruenz von Fremd- und Selbsturteil betrachtet wird. In der Regel zeigen sich in Bezug auf das BigFive-Konstrukt mittlere Korrelationen von 0,30-0,60 zwischen diversen Fremdberichten und Selbstberichten (vgl. Wirtz 2014, Vazire/Carlson 2010, Rauthmann 2017). Angesichts der hohen Korrelationen zwischen Selbst- und Metabericht verwundert es nicht, dass sich die Korrelationen (mit Zusammenhangsmaßen um 0,40) zwischen Fremd- und Metawahrnehmung ähnlich abbilden wie die zwischen Fremd- und Selbstbericht. Demnach können Personen ein bisschen ahnen, wie andere sie wahrnehmen und dass sie dies in anderer Weise tun als sie selbst, aber sie können die Fremdwahrnehmung bei weitem nicht genau einschätzen (vgl. Vazire/Carlson 2010). Beim Vergleich von Selbst- und Fremdberichten wird in gängigen Forschungen nach wie vor der Annahme gefolgt, dass hohe Übereinstimmungen eine erfolgreiche Selbstpräsentation belegen (Baumeister 1992, DePaulo 1992, Schaffhuser/Allemand/Martin 2014, Rauthmann 2017). In diesem Zusammenhang sind die Korrelationen unter zwei Aspekten zu interpretieren:

1. Es ist zu betonen, dass die Höhe der Korrelationen zwischen den Perspektiven je nach Dimension unterschiedlich ausfallen: je sichtbarer ein Merkmal ist, desto kongruenter zeichnen sich die Perspektiven ab (Back/Vazire 2012, Watson/Hubbert/Wiese 2000). Manche Aspekte sind – je nach Beobachtbarkeit und Sichtbarkeit des Merkmals – für andere, manche für den/diejenige/n selbst besser einsehbar. Entsprechend zeigen sich vor allem in den Dimensionen *Extraversion*, *Gewissenhaftigkeit* und *Offenheit für neue Erfahrungen* signifikante Korrelationen (0,20-0,65) zwischen Selbst- und Fremdberichten.⁶³ In Bezug auf die Dimension *Neurotizismus* zeigt sich dagegen, dass Selbsteinschätzungen dieser Dimension eher Teil des eigenen *Neurotizismus* sind, als dass sie tatsächlich Ausdruck in spezifischen sozialen Interaktionen finden. Zumal der *Neurotizismus* als eher schlecht zu beobachtende Eigenschaft gilt (Asendorpf/Neyer 2012). So ist

⁶³ Die entsprechende Studie verfolgte den sogenannten Settingansatz, in dessen Rahmen anhand von kurzen Videoausschnitten, Kurzgeschichten, dem Musikgeschmack oder physischem Erscheinungsbild, Website-Profilen, Schlafzimmern etc. (Back/Vazire 2012) Rückschlüsse auf Persönlichkeit gezogen wurden. So konnte gezeigt werden, dass die Persönlichkeit Spuren in der (Gestaltung der) alltäglichen Umwelt hinterlässt sowie, dass Settings persönlichkeitsabhängig bewertet und definiert sind, mit der möglichen Folge der Fehlinterpretation entsprechender Korrelationen (Asendorpf/Neyer 2012).

z.B. der *Neurotizismus* nicht gut sichtbar und gilt dementsprechend in der Selbsteinschätzung als am akkuratesten erfassbar, *Extraversion* und *Offenheit für neue Erfahrungen* demgegenüber sind gut sichtbar und können sowohl selbst als auch von nahen und entfernten Bekannten eingeschätzt werden⁶⁴ (Vazire 2010).

2. Außerdem zeigen sich bei vertieften Bekanntschaften höhere Korrelationen mit den Selbstberichten. Die Beurteilungsqualität (im Sinne einer möglichst weitreichenden Übereinstimmung von Selbst- und Bekanntenbeurteilungen) bei Eigenschaftsbeurteilungen im Rahmen von Bekanntschaften ist also abhängig von der Tiefe im Sinne von Dauer und Kontext (und damit einhergehender Informiertheit über das Verhalten) der Bekanntschaft sowie von der Art der beurteilten Eigenschaft (Asendorpf/Neyer 2012, Back/Vazire 2012). Merkmalseinschätzungen basieren demzufolge auf systematischer Verwendung vorhandener Informationen, wobei bei weniger sichtbaren Merkmalen möglicherweise auf der Selbsteinschätzung basierende Rückschlüsse gezogen werden (Watson/Hubbert/Wiese 2000).

Das eher schlechter zu beobachtende Persönlichkeitsmerkmal *Neurotizismus* zeigt demnach die deutlichste Steigerung der Korrelationswerte mit dem Grad der Bekanntschaft: so wiesen bspw. Simms/Zelazny/Yam/Gros (2010) in Bezug auf den *Neurotizismus* bei weniger als einem Jahr Bekanntschaft Korrelationen in Höhe von 0,27, bei über einem Jahr Bekanntschaft hingegen Korrelationen in Höhe von 0,46 nach (vgl. Asendorpf/Neyer 2012, dieselbe Tendenz findet sich bei Rauthmann 2017, Vazire 2010, Connelly/Ones 2010). In Bezug auf *Gewissenhaftigkeit* lassen sich demgegenüber bei naher Bekanntschaft Korrelationswerte von 0,38-0,42 ausmachen (Rauthmann 2017, Back/Vazire 2012) und in Bezug auf *Extraversion* von 0,40-0,54 (Rauthmann 2017, Back/Vazire 2012), die Werte der weiteren Dimensionen befinden sich in der Höhe zwischen 0,40 und 0,49 (Back/Vazire 2012). Connelly und Ones (2010) berichten zudem von metaanalytisch zusammengefassten Korrelationen zwischen Selbsteinschätzungen und solchen durch Fremde über alle fünf Dimensionen in Höhe von 0,08-0,22; in

⁶⁴ Das SOKA-Modell von Vazire (2010) soll bspw. darüber aufklären, welche Persönlichkeitsaspekte besser selbst und welche besser von anderen eingeschätzt werden.

der Familie hingegen von Korrelationswerten von 0,37-0,48. Andere Studien berichten von Korrelationen zwischen Selbst- und Fremdeinschätzungen der Big-Five-Dimensionen bei Verheirateten, unverheirateten Paaren und bei Freunden in Höhe von 0,03-0,36, die ebenfalls mit dem Grad der Bekanntheit und je nach Dimension variieren (allerdings zeigten sich bei dieser Untersuchung insgesamt nur eher schwache Korrelationen: von 15 Korrelationen lagen nur drei über 0,20 und insgesamt nur knapp die Hälfte über 0,10) (Watson/Hubbert/Wiese 2000).

Da Korrelationsmatrizen zwischen fremd- und selbstwahrgenommenen Persönlichkeitsmerkmalen jedoch niemals eine absolute Übereinstimmung zeigen, lässt sich schlussfolgern, dass sich Intention und Ausdruck niemals entsprechen (De Paulo 1992). Demnach gelten beide Perspektiven als valide (Hofstee 1994, Vazire 2010), wobei Fremdbenurteilungen tatsächliches Verhalten in mancher Hinsicht zuverlässiger prognostizieren als Selbstberichte (Back/Vazire 2012). Mit der Annahme, dass es nicht die eine ideale Erfassung von Persönlichkeit gibt, sondern jede Perspektive über besondere Stärken als auch Schwächen verfügt, erscheint ein Einbezug verschiedener Perspektiven die Persönlichkeit am präzisesten zu erfassen (Schaffhuser/Allemand/Martin 2014, Back/Vazire 2012).

Darüber hinaus lässt sich feststellen, dass mit steigender Abweichung des Fremdbilds vom Selbstbild dem Fremdbild verstärkt misstraut wird (Back/Vazire 2012). Generell wird sogenannten Experten bzgl. der eigenen Person eher geglaubt als anderen und soziales Feedback wird eher ins Selbstbild integriert, wenn es von jemandem kommt, der als smart und extravertiert wahrgenommen wird. Außerdem werden bewertbare Eigenschaften eher durch soziales Feedback geändert als solche, deren objektive Gegebenheit nicht von außen wahrgenommen/bewertet werden kann (Srivastava 2012).

3.4.2 Kontextspezifität

Obwohl Persönlichkeitsmerkmale als zeitlich relativ stabil gelten, zeigen sie sich situationsspezifischer, als lange Zeit angenommen wurde. Das heißt, stabile Unterschiede des Fühlens, Denkens und Handelns, durch die sich Individuen in verschiedenen Situationen und Kontexten unterscheiden lassen, gründen auf deren Persönlichkeit. Man kann also davon ausgehen, dass sich persönliches Verhalten am jeweiligen Kontext ausrichtet. Persönlichkeit ist also immer kontextgebunden zu betrachten, wobei bspw. soziale Beziehungen einen solchen übergeordneten Kontext darstellen (Schaffhuser 2014, Lievens/DeCorte/Schollaert 2008).

Der Referenzrahmen bei der Persönlichkeitserhebung reduziert dann Inkonsistenzen bzgl. ein und derselben Person (*Within Person Inconsistency*), jedoch kaum Inkonsistenzen zwischen verschiedenen Personen (*Between Person Variability*), was angesichts des fokussierten Referenzrahmens zu erwarten gewesen wäre. Diese Reduktion der *Within Person Inconsistency* zeigt sich umso deutlicher, je höher die Relevanz des Referenzrahmens für ein bestimmtes Kriterium ist. Bei allgemeinen, kontextunabhängigen Instrumenten zur Persönlichkeitserhebung zeigen sich demnach eine schlechtere Reliabilität und eine schwächere Validität (Lievens/DeCorte/Schollaert 2008), während der Einsatz eines Referenzrahmens das Antwortverhalten sowohl positiver als auch akkurater ausfallen lässt, vor allem in Bezug auf die Dimension *Gewissenhaftigkeit* (Schmitt/Ryan/Stierwalt/Powell 1995).

Allerdings ließen sich bei Holtz (2005) keine Effekte des Referenzrahmens auf die Wahrnehmung persönlicher Aspekte von Lehrkräften durch die Schülerinnen und Schüler feststellen, während ansonsten die Job-Performance als eines der relevanten Kontextkriterien für die Wahrnehmung anderer gilt (Lievens/DeCorte/Schollaert 2008).

3.4.3 Weitere Forschungsbefunde

Über den Referenzrahmen hinaus lassen sich im Längsschnitt auch altersgemäße Veränderungen innerhalb der Persönlichkeit feststellen, was für das grundsätzliche Potenzial der Wandelbarkeit von Persönlichkeitsmerkmalen spricht (Specht/Egloff/Schmukle 2011). Allerdings fallen entsprechende Ergebnisse eher uneinheitlich aus: von dem Befund von Costa und McCrae (1988), dass sich Merkmale vor allem bis zum 30. Lebensjahr verändern, bis hin zu dem Befund, dass sich alle Merkmale auch noch nach dem 30. Lebensjahr verändern (Srivastava/John/Gosling/Potter 2003). Die beiden konträren Befunde lassen sich anhand der Präzisierung erklären, nach der sich alle Merkmale bis zum 50. Lebensjahr mehr und mehr stabilisieren, danach aber wieder ausgeprägtere Persönlichkeitsveränderungen stattfinden (Roberts/DelVecchio 2000). Zudem zeigte sich, dass sich Merkmale innerhalb der Dimensionen *Extraversion* und *Neurotizismus* gleichermaßen sowohl vor als auch nach dem 30. Lebensjahr verändern können (Scollon/Diener 2006). Insgesamt ließ sich außerdem feststellen, dass Merkmale der Dimensionen *Offenheit* und *Extraversion* mit steigendem Alter zurückgehen, während *Verträglichkeit* und *Gewissenhaftigkeit* ansteigen. Die *Resilienz* als positives Pendant zum *Neurotizismus* dagegen schwankt: Bis zum 30. Lebensjahr

steigt sie in der Regel kontinuierlich an, sinkt danach, bis sie im Alter zwischen 60 und 70 Jahren am niedrigsten ist, und nimmt danach wieder zu. Hinsichtlich der Dimensionen *Offenheit*, *Gewissenhaftigkeit* und *Verträglichkeit* dagegen lassen sich jedoch nur geringe Ausprägungen der Veränderung in der Altersspanne zwischen 40 und 70 Jahren verzeichnen (Specht/Egloff/Schmukle 2011, Mayr 2014). Insgesamt gilt, dass Ältere verträglicher, gewissenhafter, resilienter sowie weniger extravertiert und weniger offen werden (Backhaus 2004). Darüber hinaus konstatieren Mayr und Neuweg (2006) anhand ihrer Forschungsbefunde, dass sich im Schnitt 50% der Varianzaufklärung (variiert leicht je nach Dimension) auf die Vererbung zurückführen lassen⁶⁵ und Umweltfaktoren vornehmlich in Kindheit und Jugend die Persönlichkeit beeinflussen. Nach Auffassung der Autoren erfolgt danach – im Sinne von Hollands Person-Umwelt-Passung (Holland 1997) – eher ein Aufsuchen von Umwelten, die der entwickelten Persönlichkeit entsprechen (Mayr/Neuweg 2006).

In Bezug auf das Geschlecht zeigen sich keine Unterschiede hinsichtlich der Veränderung von Eigenschaften, die Mittelwerte verändern sich bei beiden Geschlechtern über das Alter hinweg gleichermaßen (Specht/Egloff/Schmukle 2011). Allerdings bestehen grundsätzliche, wenn auch geringe Geschlechterunterschiede im Antwortverhalten: so geben Frauen generell höhere Werte in den Dimensionen *Neurotizismus*, *Verträglichkeit* und *Extraversion* an, keine Unterschiede ließen sich in den Dimensionen *Gewissenhaftigkeit* und *Offenheit* verzeichnen (Specht/Egloff/Schmukle 2011).

Persönlichkeitsmerkmale werden außerdem hinsichtlich ihrer Prädiktionskraft für die allgemeine Lebenszufriedenheit herangezogen. Sie gelten in diesem Zusammenhang als innere Faktoren, die äußeren Faktoren wie sozio-demographischen Merkmalen gegenüber gestellt werden. Hierbei lassen sich in mehreren Studien die Merkmale der Dimensionen *Extraversion* und *Neurotizismus* als prädiktiv wirksam ausmachen (vgl. Haller/Müller 2006).

⁶⁵ Dieser Befund resultiert aus Zwillingsstudien (siehe hierzu auch Rauthmann 2017, Asendorpf/Neyer 2012).

4 Diskurs über normative Ausrichtungen der Unterrichtsqualität

Der Begriff *Qualität*, abgeleitet vom lateinischen *qualitas*, bezieht sich im engeren Sinne zunächst einmal nur auf die reine Beschaffenheit oder Eigenschaft einer Sache oder Person. Erst im weiteren Verständnis bezieht sich der Begriff auf die Güte dieser Eigenschaften von Personen oder der Beschaffenheit von Sachen als „einschätzende, objektivierte Bewertung der Güte, des Wertes oder des allgemeinen Niveaus eines Objekts (im weiteren Sinne)“ (Terhart 2000, S. 814). Entsprechend lässt sich die Definition des Qualitätsbegriffs in Bezug auf Normung und Standardisierung von Produkten und Abläufen ableiten: „Qualität ist die Beschaffenheit einer Einheit bezüglich der Eignung, festgelegte und vorausgesetzte Erfordernisse zu erfüllen“ (DIN 553550, Teil 11, in Terhart 2000, S. 815). Im Bildungsbereich wird die Normung und Standardisierung des Qualitätsbegriffs

„auf der Basis von normativ gesetzten und begründeten Zwecken und Zielbeschreibungen definiert. [...] Die Quellen für solche normativen Setzungen können Traditionen, allgemein- und sozialpolitische Überzeugungen, sozial- und bildungstheoretische Erwägungen und Entscheidungen, können normative Menschenbildannahmen sowie globale Zielvorstellungen über Personen-, Gesellschaft- oder gar Menschheitsentwicklung generell sein.“ (Terhart 2000, S. 815f)

Wenn es diesbezüglich aufgrund fehlender einheitlicher Maßstäbe keine einheitliche Definition geben kann und sich eine entsprechende inhaltliche Bedeutung jedoch erst im Kontext herauszubilden vermag (vgl. Heid 2000, Harvey/Green 2000, Gräsel/Göbel 2015), müssen Konzepte der *Qualität* stets neu ausgehandelt werden (Klieme 2013). Festhalten lässt sich, dass *Qualität* über eine positiv besetzte Wortbedeutung verfügt, denn „wer kann und will schon [...] gegen Qualität sein?“ (Terhart 2000, S. 809), wobei Ball schon 1985 fragte: „Was um alles in der Welt ist Qualität?“ (Ball in Harvey/Green 2000, S. 18)

In Bezug auf die *Qualität* im Bildungsbereich stellt Heid (2000) fest, dass diese nicht lediglich die „beobachtbare Eigenschaft oder Beschaffenheit eines Objekts, sondern das Resultat einer Bewertung der Beschaffenheit eines Objekts“ (S. 41) darstellt. Das heißt, dass *Qualität* hier keinen Sachverhalt an sich abbildet, der

außerhalb wertender Stellungnahmen existiert, sondern *Qualität* immer erst zugeschrieben wird. Damit ergibt sich das Problem der Parallelität wertender Stellungnahmen (siehe auch Gräsel/Göbel 2015). Es erscheint zunächst keine Position legitimer als andere (Heid 2000), da Qualitätsfeststellungen immer nur auf normativen Grundlagen basieren, was auf die Frage nach der Legitimation der jeweiligen Feststellung statt auf deren Richtigkeit oder Wahrheitscharakter hinausläuft (Klieme/Tippelt 2008). Folgerichtig kann aus dieser Perspektive zum einen argumentiert werden, dass die *gute* Schule ein Ergebnis tautologischer Argumentation ist (vgl. Heid 2000), und zum anderen, dass kein Konsens darüber, was eine gute Schule ist, bestehen kann, da es ein ungeklärtes Problem bleibt,

„auf welcher Ebene (Schüler und Lehrer, Unterricht, Schule, Schulsystem), mit welchen Verfahren (qualitativ vs. quantitativ; intern vs. extern), anhand welcher Bewertungsmaßstäbe und Erwartungen (ökonomisch vs. pädagogische) reliable und valide Aussagen über die schulische Qualität oder Grade der Zielerreichung getroffen werden können.“ (Ditton, 2000, S. 75)

Damit wird offenbar, dass nicht die eigentlichen Resultate qualitätsbezogener Untersuchungen problematisch sind, sondern vielmehr die Legitimation und die Transparenz der dahinterstehenden Werthaltungen und ihrer Kriterien (Heid 2000).

Nachdem in der Qualitätssicherung im Bildungsbereich zunächst die Inputsteuerung im Mittelpunkt stand, womit Ressourcen und Bedingungen in das Blickfeld rückten, folgte darauf in logischer Konsequenz der output- bzw. produktorientierte Blick auf die entsprechenden Effekte im Sinne kurzfristiger Outputs sowie längerfristiger Outcomes (Klieme/Tippelt 2008). Effektivität kann dabei entlang des Grad des Erreichens anvisierter direkter und indirekter Ziele in Form von *Outcomes* definiert werden (Harris/Rutledge 2007); Lehreffektivität dementsprechend als Grad der Vermittlung von Erziehungs- und Bildungszielen, die sich in Form von Schülerleistungen – als *Intermediate Outcomes* – offenbaren (McKeachie 1979). Darauf folgte die bis heute weitestgehend gängige Prozessorientierung, die die Bildungsqualität anhand ihrer Prozessqualität im direkten Unterrichtsverlauf definiert (Klieme/Tippelt 2008). Dieser Ansatz bündelt Lehr-Lern-Prozesse entlang eines *Angebot-Nutzungs-Modells* (Klieme/Tippelt 2008, Helmke 2004, 2007, 2014).

Hierbei sind in den Arbeiten zur Theorieentwicklung und Modellbildung der Qualitätsforschung grundsätzlich zwei Dimensionen zu unterscheiden: zum ei-

nen die strukturelle Dimension, die Schule grundsätzlich als ein Mehrebenensystem verortet, und zum anderen die dynamische Dimension, die die „[...] Prozesse im Schulwesen, durch die Eingangsbedingungen (inputs) in erzielte Ergebnisse (output, outcomes) transformiert werden“, betrachtet (Ditton 2000, S. 76). Hinsichtlich der strukturellen Dimension zeigte sich in den Schulsystemvergleichen der 1970er Jahre, dass innerhalb desselben Schulsystems immense Unterschiede zwischen den Einzelschulen bestehen können (vgl. Fend 1977). Demnach muss auf der Ebene der (Einzel)Schule das Konzept der Qualität stets neu ausgehandelt werden (vgl. Terhart 2000). Eine einheitlich festgelegte Definition bringt nämlich unlösbar konzeptuelle Probleme mit sich, da die Urteile über die Schulqualität einer normativen Basis, d.h. legitimierbarer Gütekriterien bedürfen, die ihrerseits wiederum von politischen und pädagogischen Interessen abhängen (Klieme 2013). Aus empirischer Forschungsperspektive bedeutet das, dass mit der erklärten Gültigkeit einer (Schul-)Eigenschaft als Gütekriterium der wertend-normative Prozess erst seinen Anfang nimmt (Klieme 2013). Auf dieser argumentativen Grundlage lässt sich die zunehmende Verzahnung der beiden Dimensionen *strukturell* und *dynamisch* (Ditton 2000) entlang des Wechsels vom Bereich der Schulsystemanalyse, die sich auf den Nachweis von Unterschieden innerhalb gleicher Systeme konzentriert, hin zu der Qualität der Erfahrungen von Betroffenen vor Ort erklären (vgl. Terhart 2000). Dabei wird Qualität anhand der Ebene des konkreten Handlungsverlaufs definiert (Terhart 2000), was sich dann auf den jeweiligen Unterricht bezieht. Entsprechend fordert Ditton (2000), dass sich Forschung zu „schulischer Qualität bzw. Effektivität primär auf die Ebene des Unterrichts beziehen [muss] – in der Regel auf die Ebene der einzelnen Schulklasse“ (S. 73). Zumal er außerdem damit argumentiert, dass bisherige Forschungsergebnisse den Schluss nahelegen, dass sich Wirkungen nicht einheitlich auf der Schulebene ausmachen lassen, sondern die Unterschiede zwischen den Klassen primär der einzelnen Lehrkraft und ihrem Unterricht zuzuordnen sind. Heute gilt es als weitgehender Konsens, dass Effekte vor allem bei schülernahen Bedingungsfaktoren innerhalb von Lehr-Lernsituationen respektive Prozessmerkmalen des Unterrichts auftreten (Ditton 2000).

In diesem Zuge unternimmt Klieme (2013) den Versuch, Prozessmerkmale des Unterrichts weitgehend aus der Qualitätsdebatte herauszunehmen, indem er feststellt, dass empirisch überprüfbare Prozesse zunächst einmal nur beobachtbar sind, ohne eine Wertzuschreibung zu implizieren. Das heißt, es werden zuerst

Merkmalskataloge angeboten und erst, wenn ein Zusammenhang mit Lernerfolgen nachgewiesen ist, gelten solche Prozessmerkmale als Qualitätsmerkmale. Dadurch bleiben diese Merkmale nur im weitesten Sinne mit dem Qualitätsbegriff verknüpft und gehen mit dem heutigen Erkenntnisstand weit über eine rein begriffsanalytische Betrachtung des Qualitätsbegriffs hinaus (Klieme 2013). Eine entsprechende Theorie mit inhaltlichen Abgrenzungen und theoretischen Ordnungssystemen erscheint allerdings eher mangelhaft, da sie keinen Konsens bezüglich relevanter übergeordneter Unterrichtsmerkmale zu schaffen vermag, sondern insgesamt uneinheitlich und unvollständig auftritt (Schröder 1999). Dieser Umstand mag darin begründet sein, dass sich vielfältige Traditionen und Paradigmen konfliktuell gegenüberstehen (Klieme/Tippelt 2008).

Hier ließe sich erneut mit der Relationalität des Qualitätsbegriffs ansetzen, nämlich mit der Frage nach der Perspektive der Qualitätsbeurteilung, die sich je nach Denkweise, philosophischer, politischer Grundlage und Weltbild unterscheidet, jenseits von richtig oder falsch (Harvey/Green 2000). Alle Einschätzungen dieser Art beziehen sich innerhalb des Bildungsbereichs jedoch implizit auf Kriterien des Prozesses und des Produkts. Dabei soll der Prozess zum erwünschten Produkt führen und das Produkt soll wiederum seine funktionale Zweckmäßigkeit erfüllen, die allerdings je nach Interessengruppe variieren kann. Neben einer solchen Qualitätskategorie der *Zweckmäßigkeit* ist zudem noch die Kategorie der *Perfektion* bedeutsam, die die präventive Fehlerlosigkeit, Qualitätskultur und Prozessspezifikation statt im Gegenteil dazu eine Qualitätskontrolle im Sinne von Input-/Outputsteuerung (vgl. S. 22) anvisiert, um letztlich Perfektion und Zweckmäßigkeit zu einer Transformation zu verknüpfen. Diese Transformation sehen Harvey und Green (2000) in der Weiterentwicklung der Lernenden, die durch eine Machtverlagerung der Definitionshoheit zu mehr Selbstbestimmung hinführt. So gesehen erscheint es als „Zeitverschwendung, Qualität einheitlich definieren zu wollen“ (Vroeijensteyn 1991, in Harvey/Green 2000, S. 36). Bestenfalls ließe sich präzise definieren, welche Kriterien verschiedene Interessengruppen bei der Qualitätsbeurteilung anwenden und aus welchen Perspektiven Qualität eingeschätzt wird (vgl. Harvey/Green 2000).

In diesem Sinne lässt sich Unterrichtsqualität in konstruktivistischer Manier als Kombination aus Normativität hinsichtlich der Zweckmäßigkeit des Lernerfolgs und Individualisierung hinsichtlich der Perfektion sowie der Transformation aus Schülerperspektive begreifen. Zwar bleiben die Ziele des Unterrichts und der

solchermaßen definierte Lernerfolg normativ gesetzt, und zudem ist der Zusammenhang zwischen Unterrichtsprozessen und Lernerfolgen vor allem empirisch begründet (Clausen/Reusser/Klieme 2003); allerdings kann die Qualitätsbeurteilung der Prozessmerkmale durch die Schülerinnen und Schüler selbst erfolgen, womit deren Perspektive wesentliche Berücksichtigung findet. Vor dem Hintergrund dieses gedanklichen Konstrukts kann das Angebot-Nutzungs-Modell verstanden werden, in dem Unterricht als ein Angebot von Lerngelegenheiten, die von den Schülerinnen und Schülern individuell wahrgenommen und genutzt werden, gilt (Rakoczy/Buff/Lipowsky 2005).

Abschließend wird festgehalten, dass in dieser Arbeit Unterrichtsqualität zunächst als Beschaffenheit von Prozessmerkmalen aus Schülerperspektive, die je nach Art der Unterrichtsbeteiligung und gemäß der jeweiligen Perspektive unterschiedlich (vgl. auch Clausen 2002) positiv oder negativ ausgeprägt sein kann, verstanden wird.

4.1 Forschungslinien und Forschungstraditionen zur Unterrichtsqualität

Ähnlich dem Konstrukt der *Persönlichkeit* liegen auch der *Unterrichtsqualität* je nach Tradition und Ausrichtung der Unterrichtsforschung unterschiedliche Prinzipien zugrunde, die kaum miteinander vernetzt sind; diese reichen von der Lehr-Lernforschung über die *Instructional Design*-Forschung bis zur Klimaforschung (Helmke 2009). Clausen (2002) klassifiziert die verschiedenen theoretischen Ansätze in der Forschung zur Unterrichtsqualität folgendermaßen:

- a) Pädagogische Konzepte: guter Unterricht beinhaltet normative Aspekte hinsichtlich der verschiedenen Ziele, die dem Unterricht zugeordnet werden können wie Effektivität (Prozess-Produkt-Paradigma), kognitive Aktivierung (Konstruktivismus), Förderung des Selbstkonzepts (Motivationspsychologie), Förderung des sozialen Lernens (Reformpädagogik), Divergenzminderung in der Klasse (Systemtheorie)
- b) Empirische Forschung: vermehrte Berücksichtigung der bidirektionalen Interaktion und bidirektionalen Wirkungen von Verhalten und Wahrnehmung zwischen Lehrenden und Lernenden mit Fokus auf die aktive Rolle der Lernenden; Vereinigung der Ziele *Leistung* und *psychosozialer Entwicklung/Motivation*

- c) Unterrichtsklimaforschung: Erfassung der Schülerperspektiven/-wahrnehmungen
- d) Lehr-Lern-Forschung: Fokus auf verhaltensbezogene Unterrichts-/Prozess- und Lehrermerkmale (ähnlich zu Prozess-Produkt-Paradigma)

Jedoch konnten Clausen (2002) und Gruehn (2000) nachweisen, dass sich verschiedene Forschungsrichtungen durchaus kombinieren lassen. So lassen sich Ergebnisse aus der Verknüpfung von Erkenntnissen der Lehr-Lernforschung mit jenen der Klassenklimaforschung als *valide* bezeichnen, womit sie für eine erweiterte Beachtung der Schülerperspektiven auf Klassenebene innerhalb der Lehr-Lernforschung sprechen (vgl. Gruehn 2000, Clausen 2002). Nach Clausen (2002) geht die reine Unterrichtsqualität als *Metakonstrukt* aufgrund konkurrierender Zielgerichtetheit (je nach pädagogischer Theorie) jedoch über die reine Interaktionsebene hinaus. Wird die Qualität von Unterricht dementsprechend als Beschaffenheit eines Sachverhalts, die je nach Perspektive positiv oder negativ ausgeprägt sein kann, verstanden, lassen sich die verschiedenen Strömungen in drei große Richtungen einteilen: die Personenorientierung, die Produktorientierung und die Prozessorientierung. Innerhalb dieser Richtungen lassen sich verschiedene Forschungslinien mit unterschiedlichen Schwerpunkten einordnen. Die ersten Versuche in der Frühzeit der Unterrichtsforschung vor mehr als hundert Jahren folgten zunächst vor allem der Personenorientierung. Dabei richteten sich die Forschungen auf die Suche nach dem *guten* Lehrer resp. der *guten* Lehrerin anhand individueller Persönlichkeitsmerkmale aus. Angesichts der völligen Ausblendung des tatsächlichen Unterrichtsgeschehens blieb dieses Vorgehen allerdings weitgehend erfolglos. In den 50er und 60er Jahren des zwanzigsten Jahrhunderts entsprach die empirische Unterrichtsforschung vornehmlich der amerikanisch orientierten *Teacher Effectiveness Research*, die sich gemäß des Persönlichkeitsparadigmas auf die Offenlegung von persönlichen Merkmalen erfolgreicher Lehrer im Sinne eines hohen Leistungsniveaus bzw. einer positiven Leistungsentwicklung seitens der unterrichteten Schülerinnen und Schüler fokussierte (Helmke 2009). Konkret wurden dabei Einstellungen und Kenntnisse der Lehrperson mit dem Produkt der Schülerleistungen – weitgehend ohne Erfolg – verknüpft, wobei das Unterrichtshandeln als solches ausgeblendet blieb (Blömeke 2009). Angesichts dessen stagnierte die Forschung unter diesem Vorgehen, da sich nur vereinzelt eindeutige Zusammenhänge von Merkmalen der Lehrerpersönlichkeit mit den Leistungen der Schülerinnen und Schüler nachweisen ließen:

beispielsweise der Humor, das Engagement und das Selbstwirksamkeitsempfinden von Lehrkräften. Diese Befunde werden jedoch heute im Zuge von Professionalisierungsdebatten entlang des Expertenparadigmas, das die Forschung zur Lehrerpersönlichkeit weitgehend abgelöst hat, neu eingeordnet (Helmke 2009, S. 46). In den nächsten Jahrzehnten, den 1970ern und 1980ern orientierte sich die empirische Unterrichtsforschung vermehrt an anderen Forschungsrichtungen, vor allem dem Prozess-Produkt-Paradigma (Helmke 2009), innerhalb dessen vermehrt der Einzelschüler in den Blick genommen wurde, während die Bedeutung der Lehrerpersönlichkeit immer weiter in den Hintergrund trat (Blömeke 2009). Das Prozess-Produkt-Paradigma gilt Helmke zufolge als *klassisches* Paradigma, dessen Forschung generell auf drei Vorgehensschritten gründet: 1. der Erhebung von Prozessmerkmalen des Unterrichts, 2. der Messung von Zielkriterien, die in der Regel Lernleistungen der Schülerinnen und Schüler abbilden und 3. den Berechnungen von Zusammenhängen (Korrelationen) zwischen den erhobenen Prozessmerkmalen und Zielkriterien. In dieses Forschungsvorgehen wurden auch Strömungen wie das Expertenparadigma oder experimentelle Trainingsstudien implementiert (Helmke 2009). Das Modell wurde nach und nach modifiziert, etwa indem Wechselwirkungen von Unterrichtsqualität mit schulischen und die Evaluation beeinflussenden Rahmenbedingungen des Unterrichts und innerhalb der Schülerinnen und Schüler liegende Mediationsprozesse sowie Mehrebenenmodelle berücksichtigt wurden (Helmke 2009). Außerdem konnten einige Merkmale des Unterrichtsprozesses ausgemacht werden, die valide mit den Lernergebnissen der Schülerinnen und Schüler korrelieren.

Die Personenorientierung der Frühzeit findet sich heutzutage vornehmlich in den Ansätzen zu *Kompetenz* und *Professionalität* sowie in Bezug auf sogenannte *Assessment Centers* für Lehramtsanwärter wieder (Helmke 2006, mehr dazu in Kap. 2.2.4). Dabei besteht bis heute ein weitreichender Konsens darüber, dass erstens schwer erlernbare, relativ stabile persönliche Eigenschaften des Lehrers (womit insbesondere Freude, Enthusiasmus, Sensibilität, Frustrationstoleranz gemeint sind) wichtige Bedingungen guten Unterrichts sind, und zweitens es vor allem subjektive Theorien, also Alltagstheorien zum Lehren und Lernen sind, die das berufliche, d.h. unterrichtliche Handeln der Lehrperson bestimmen (Helmke 2006). So konstatiert Helmke (2007), dass die Empirie zwar bloße Spekulationen ablöst, jedoch „Unterricht [...] auch heute vielfach noch als ein unantastbares Feld der freien Entfaltung der Lehrerpersönlichkeit [gilt]“ (S. 11). Allerdings

finden sich zu dieser Behauptung Helmkes keinerlei Belege, außer dass Lehrerverkooperation im Unterricht äußerst selten stattfindet, was in der DESI-Studie nachgewiesen wurde (vgl. Helmke/Schrader/Wagner et al. 2008). Darüber hinaus stehen sich generell variablenorientierte Strömungen, die jeweils eher der Prozess- oder der Produktorientierung zuzuordnen sind, gegenüber (Helmke 2006). Die Prozessorientierung bezieht sich dabei auf die konkrete Durchführung des Unterrichts, der normativ-präskriptiv erfasst wird. Dabei wird die Qualität des Unterrichts anhand von Unterrichtsmerkmalen (auch Prozessmerkmale genannt), die sich im Unterrichtshandeln und der Lehrer-Schüler-Interaktion offenbaren, erfasst (Helmke 2007). Die Produktorientierung dagegen bezieht sich auf den Effekt von Unterricht und Schule, sprich: auf *Outputs* oder *Outcomes*, d.h. auf erbrachte und nachweisbare, messbare Leistungen bzw. die Lernentwicklung der Schülerinnen und Schüler. Dabei werden die Leistungen der Schülerinnen und Schüler als Indikatoren der Unterrichtsqualität verwendet (Helmke 2007). Es gibt zwar außerdem neuere Bemühungen um die Erforschung entlang der *Best Practice*-Orientierung⁶⁶, allerdings stehen diese vor dem Problem der Beliebigkeit und Normativierung vermeintlich guten Unterrichts und sind zudem empirisch wenig erforscht (Helmke 2006), womit wieder (wie in Kap. 3.2 hinsichtlich des Konstrukts *Persönlichkeit* diskutiert) das Problem aufkommt, dass etliche Merkmalskataloge guten Unterrichts nebeneinander existieren. Die Idee des Angebot-Nutzungs-Modells von Helmke (u.a. 2004, 2007, 2009, 2014) unternimmt jedoch den Versuch alle drei großen Orientierungen (Personenorientierung, Produktorientierung, Prozessorientierung) zu integrieren (siehe Abb. 3):

⁶⁶ Hierbei wird versucht Merkmale von Unterricht, der nachweislich zu besonders guten respektive gewünschten *Outputs* führt, auszumachen. Zumal dieses Vorgehen im Sinne von *Expertenlehrern* auch wieder an der Personenorientierung ansetzt.

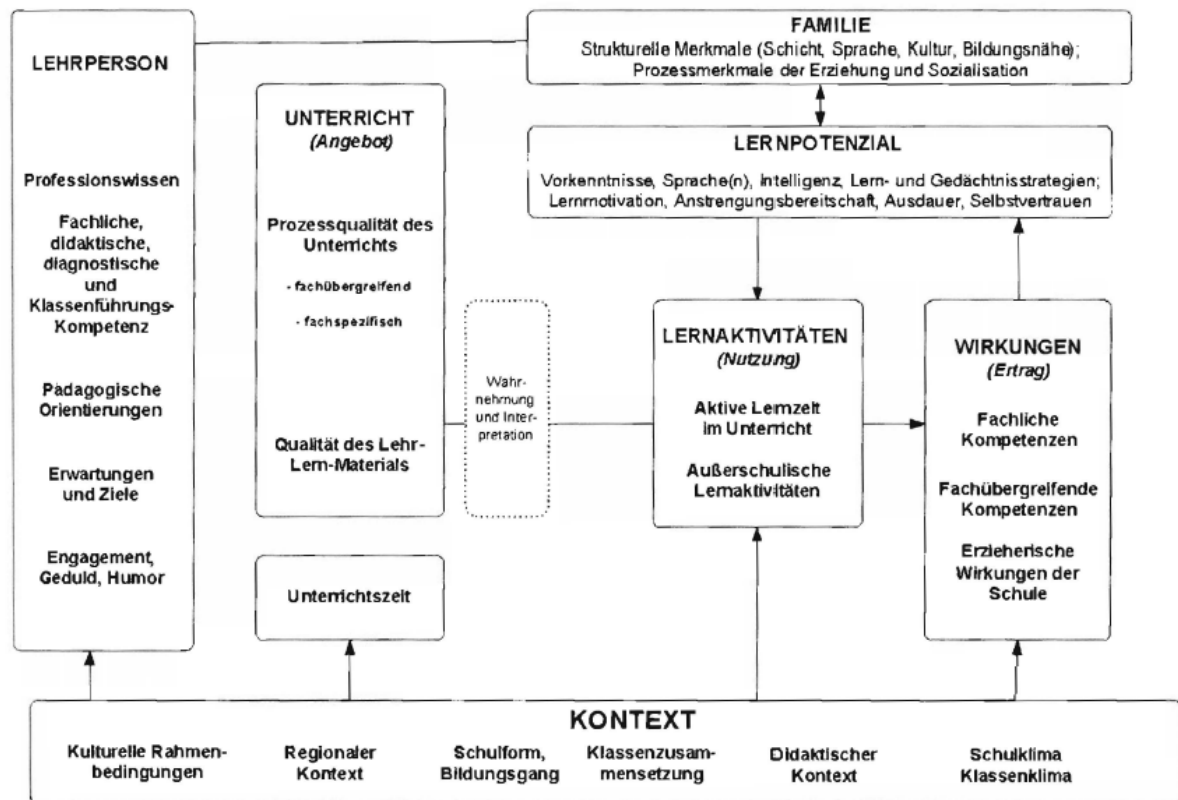


Abb. 3: Angebots-Nutzungs-Modell des Unterrichts (Helmke 2014, S. 71).

Auf Basis des Prinzips der Lehr-Lernforschung unter Beachtung konstruktivistischer Klimaansätze und mit besonderem Fokus auf die Kombination von Prozess- und Produktparadigma (zusammengefasst als Prozess-Produkt-Paradigma) bildet dieses Modell Zusammenhänge zwischen Prozessmerkmalen des Unterrichts (als Lehrer-Schüler-Interaktion) und dem Produkt des Unterrichts (mit der Lernleistung als Produktmaß) ab. Dabei werden bestimmte Wirkungen des Unterrichts sowie deren Wechselwirkungen mit Schülermerkmalen sowie diesbezügliche empirische Befunde impliziert. Diese allerdings bis dato noch immer wenig untersuchten Wirkungen und Wechselwirkungen entspringen diversen Forschungen zu Unterrichtsmerkmalen, die unabhängig voneinander und trotz unterschiedlicher Perspektiven zu ähnlichen und bestenfalls fachübergreifend validen Ergebnissen gelangen (Helmke 2006, 2009).

4.2 Operationalisierung des Konstrukts *Unterrichtsqualität*

Während sich Unterricht als „didaktisch geplante und deshalb thematisch abgrenzbare als auch zeitlich hinreichend umfassende Sequenzen des Lehrens und Lernens im Kontext pädagogischer Institutionen“ (Helmke 2009, S. 44) definieren

lässt, bestimmt die empirische Unterrichtsforschung *guten* Unterricht vornehmlich als *erfolgreichen* Unterricht im Sinne der Wirkungen und Effekte auf die Lernentwicklung der Schülerinnen und Schüler (Arnold 2007). In diesem Sinne wird zwar versucht, den Begriff der Unterrichtsqualität rein funktional zu prägen und normative Aspekte außen vor zu lassen (Clausen/Klieme/Reusser 2003); allerdings lässt sich die Lernleistung von Schülerinnen und Schülern ebenso als Norm ansehen wie andere Bewertungsperspektiven auch, zumal, wie sich im Weiteren zeigen wird, *guter* Unterricht nicht allein an den Lernergebnissen der Schülerinnen und Schüler festzumachen ist, da solche Lernergebnisse immer auch in weiten Teilen von anderen Faktoren wie individuellen Schülermerkmalen und schulischen Rahmenbedingungen abhängen (Jones 1989, Greimel 2002).

Empirische Bewertungsperspektiven auf die Unterrichtsqualität – sowohl in Bezug auf Unterrichtsprozesse als auch auf Unterrichtsprodukte, wie es z.B. bei Erhebungen wie PISA, TIMSS und der Einführung von Bildungsstandards der Fall ist – können entlang unterschiedlicher Bezugsnormen erfolgen: entweder vergleichend, kriterial oder veränderungsbezogen (Helmke 2009, S. 47f). Angesichts der relativen Uneinheitlichkeit der Definition von Unterrichtsqualität ist eine Manifestation dieser Bezugsnormen allerdings schwierig (Helmke 2009). Clausen, Klieme und Reusser (2003) verorten die Bezugsnormen daher wiederum anders, und zwar entlang der herbartianischen, der reformpädagogischen und der fachdidaktischen Tradition.⁶⁷ Es wird dabei deutlich, dass der Unterrichtsforschung je nach Forschungsparadigma oder -tradition unterschiedliche Prinzipien zugrunde liegen, die jedoch kaum miteinander verknüpft sind (vgl. auch Helmke 2009).

Zu den verschiedenen Ansätzen kommt der Umstand hinzu, dass Unterrichtsqualität zudem aus verschiedenen Perspektiven erhoben werden kann, die kaum miteinander zusammenhängen (vgl. Clausen 2002, Fauth et al. 2014): aus Lehrerperspektive, aus Schülerperspektive, aus externer Beobachterperspektive (auch per Videographie). Jeder der Perspektiven können verschiedene Vor- und Nachteile zugesprochen werden (siehe hierzu Clausen 2002, Gruehn 2000). Während

⁶⁷ Zur Ergänzung: herbartianisch, als dass die Lehrperson systematisch strukturierte Interaktionsprozesse plant und steuert; reformpädagogisch im Sinne der Selbsttätigkeit der Schülerinnen und Schüler innerhalb einer interaktiven Lernumwelt; fachdidaktisch als produktive Aneignung von Konzepten durch fachsystematisch aufbauende Lehr-Lernaktivitäten (Clausen/Klieme/Reusser 2003).

sich die Kompetenz der Lehrkraft vor allem situationsspezifisch im Unterricht zeigt, basieren bisherige Studien jedoch noch vornehmlich auf der Selbstausskunft der Lehrkraft, die beispielsweise durch die soziale Erwünschtheit oder unterschiedliche Bezugsnormorientierungen verzerrt und damit nicht als generell aussagekräftig angesehen werden können (Blömeke 2009). Vermehrt erfolgt seit den 1990er und 2000er Jahren jedoch mit der Klassenklimaforschung eine zunehmende Aufmerksamkeit auf Schülerfeedbacks sowie eine Verknüpfung der verschiedenen Perspektiven miteinander. Dabei zeigte sich, dass sich mit 30%iger Varianzaufklärung die Lehrkraft neben individuellen Schülervoraussetzungen als höchste Wirkkraft in Bezug auf die Schülerleistungen und im Gesamtgefüge *Unterrichtsqualität* verzeichnen lässt (Blömeke 2009).

„In der Empirischen Unterrichtsforschung [...] charakterisiert man Unterricht im Hinblick auf bestimmte Qualitätsdimensionen (z.B. Strukturiertheit, Verständlichkeit, Motivierung), die nachweislich (belegt durch vorangegangene Untersuchungen) eine Rolle für den Lernerfolg spielen. Es handelt sich also um eine ‚variablenorientierte‘ Betrachtungsweise, die davon ausgeht, dass der Unterricht verschiedener Lehrpersonen hinsichtlich bestimmter Merkmale (Variablen), die von Lehrkraft zu Lehrkraft, aber auch bei der gleichen Lehrkraft in verschiedenen Situationen (Fächer, Klassen, Zeitpunkte) variieren kann (z.B. hohe Ausprägung der Variable ‚Verständlichkeit‘ bei Lehrperson A, niedrige Ausprägung dieser Variable bei Lehrperson B).“ (Helmke/Schrader 2008, S. 25)

Vermehrt wurden einhergehend mit der empirischen Wende und den Ansätzen aus der Klimaforschung Schülerurteile zur Beschreibung der Unterrichtsqualität eingesetzt. Zumal aus lernpsychologischer Sicht grundsätzlich gilt, dass die Lehrkraft nicht direkt das Lernen der Schülerinnen und Schüler beeinflussen kann, sondern lediglich in der Absicht eine Umgebung schaffen kann, dass diese von den Lernenden adäquat wahrgenommen und in eigene Lernaktivität umgesetzt wird (vgl. Silbereisen 1996). Diesem Grundsatz folgen das Angebot-Nutzungs-Modell von Helmke (u.a. 2004, 2007, 2009, 2014) und Modelle von Klieme (u.a. 2006, Klieme/Lipowsky/Rakoczy/Ratzka 2006, Klieme/Rakoczy 2008). Insgesamt lässt sich allerdings konstatieren, dass die Forschung in Deutschland lange Zeit rein geisteswissenschaftlich und auch im Hinblick auf Lernzielkriterien ideell ausgerichtet war und noch immer, trotz empirischer

Wende, angesichts eines Mangels an empirischer Forschung als rückständig bezeichnet werden kann (Helmke 2007).

4.3 Prozessmerkmale der Unterrichtsqualität

Angesichts der in dieser Arbeit besonders im Fokus stehenden Wahrnehmungsperspektive und davon ausgehend, dass für den Lernerfolg von Schülerinnen und Schülern vornehmlich deren Wahrnehmung von Unterricht ausschlaggebend ist (vgl. Arnold 2007), werden hier solche Möglichkeiten der empirischen Erfassung ausgewählt, die die Perspektiven der Schülerinnen und Schüler und diese möglichst differenziert, d.h. niedrig inferent, erfassen.

Die Variation hinsichtlich der für die Lernentwicklung von Schülerinnen und Schülern als relevant erachteten Unterrichtsmerkmale ist ähnlich breit aufgestellt wie die Identifikation von Persönlichkeitsmerkmalen. Je nach Forschungsschwerpunkt sowie Zielkriterium (i.d.R. Lernleistungen) unterscheiden sich Zusammenstellung und Bedeutsamkeit der jeweiligen Merkmale. Merkmale *guten* Unterrichts können daher zwar als Leitlinien, Prinzipien oder Orientierungspunkte, nicht aber als sogenannte *Rezepte* aufgefasst werden. Zudem ist zu beachten, dass nicht immer eine maximale Ausprägung von Merkmalen gleichbedeutend mit einer hohen Unterrichtsqualität ist. So können sich teils lineare, teils u-förmige Ausprägungsverläufe als optimal erweisen (Helmke 2006).

4.3.1 Entwicklung von Prozessmerkmalen des Unterrichts

TIMSS 1995 und 1999 weisen auf einen empirischen Zusammenhang zwischen Schülerleistungen und schülerperzipierten Unterrichtsmerkmalen hin, woraus sich ein Angebots-Nutzungs-Modell entwickelte, in dem Unterricht als Angebot von Lerngelegenheiten gilt, die von den Schülerinnen und Schülern individuell wahrgenommen und genutzt werden. Um diese Annahme zu verifizieren, bezog sich die zentrale Fragestellung des Pythagoras-Projekts (einer schweizerisch-deutschen Videostudie) auf die Bedeutung unterrichtlicher und außerunterrichtlicher Variablen für die Leistungs- und Interessensentwicklung im Mathematikunterricht, was anhand eines Vergleichs deutscher Schülerinnen und Schüler mit schweizerischen Schülerinnen und Schülern, die gegenüber den deutschen einen deutlichen Leistungsvorsprung in den Leistungstests zeigten, überprüft und zumindest teilweise bestätigt werden konnte (Rakoczy/Buff/Lipowsky 2005).

Mit der Zeit haben sich verschiedene Prozessmerkmale des Unterrichts herauskristallisiert, die sich sowohl innerhalb der Prozess-Produkt-Forschung als auch der Klassenklimaforschung verorten lassen (Clausen/Reusser/Klieme 2003). So zeichnen sich Merkmale wie z.B. Strukturiertheit, Verständlichkeit und Motivierung als nachweislich relevant hinsichtlich der Lernerfolge von Schülerinnen und Schülern ab (Helmke 2007). Wobei Arnold (2007) diesen Punkt kritisiert, da er die Lernwirksamkeit lediglich als eine vermutete Folge des Unterrichtsangebots als adaptive Realisierung der Lehrplanung von Lehrerinnen und Lehrern ansieht. Allerdings zeigt sich eine deutliche Variation dieser Merkmale von Unterricht zu Unterricht, die mit einer entsprechenden Variation der Lernerfolge einhergeht⁶⁸, was für die Validität der Befunde spricht (Helmke 2007). Fachübergreifende Prozessmerkmale der Unterrichtsqualität zeichnen sich außerdem sowohl durch direkte als auch indirekte Effekte auf die Lernentwicklung der Schülerinnen und Schüler aus (Helmke 2007). Schulisches Lernen wird dabei stärker von direkten (wie Classroom Management, Student/Teacher Social Interactions, Quantity of Instruction) als von indirekten (wie Politik, Kultur, Schulform, indirekte Einflüsse tangieren höchstens direkte Einflussvariablen) Einflüssen tangiert (vgl. Wachtel 1998). Anhand des faktorenanalytischen Vorgehens der pädagogischen Psychologie über empirisch prädiktive schulleistungsrelevante Einzelmerkmale konnten Schlüsselmerkmale der Prozessqualität von Unterricht ermittelt werden. Als ein solches offenbarte sich bspw. die Klassenführung (Helmke 2006), die schon Kounin (1976) als wesentliches und übergeordnetes Merkmal für Unterrichtsqualität identifizierte. Klieme und Rakoczy (2008) dagegen teilen verschiedene Merkmale innerhalb dreier übergeordneter Unterrichtsdimensionen ein: Klassenführung, kognitive Aktivierung und Schülerorientierung.

⁶⁸ Fähigkeiten der Schülerinnen und Schüler sowie die Unterrichtsgestaltung zeigten – in einer Metaanalyse (Ergebnisse wurden dabei in einen konzeptionellen Rahmen, basierend auf Modellen schulischen Lernens, die von Einflüssen auf das Lernen ausgehen, eingeordnet (28 Unterkategorien in sechs Hauptkategorien wie Fähigkeiten des Schülers, Unterrichtsgestaltung, Kontext etc.)) über 11000 statistische Befunde aus den 40er bis 90er Jahren – insgesamt und fast gleichauf die stärksten Effekte auf Lernerfolge. Dabei zeigten sich vor allem Classroom Management, metakognitive und kognitive Prozesse des Lernens und die Schüler-Lehrer-Interaktion als bedeutsam, bezüglich der Unterrichtsgestaltung außerdem die effektive Unterrichtszeit, das Klassenklima und Techniken der Unterrichtsgestaltung (Wang/Haertel/Walberg 1993).

Insgesamt existiert eine schwer überschaubare Datenbasis zu Merkmalen der Schul- und Unterrichtsqualität, die immer wieder neu zu strukturieren versucht wird (Schröder 1999, Hattie 2013). Versuche systematischer Zusammensetzungen erfolgten u.a. von Carroll (1963, 1973, 1989), Slavin (1994) und Schröder (1999) und nicht zuletzt von Hattie (2013). Schröder (1999) konnte allerdings mit ihrer Arbeit aufzeigen, dass neben den uneinheitlichen und unvollständigen Ordnungssystemen von Unterrichtsmerkmalen in der Literatur auch die Ordnungsprinzipien von Experten deutlich variieren und kein theoretisch fundierter Konsens bzgl. qualitativ relevanter übergeordneter Unterrichtsmerkmale besteht (vgl. Schröder 1999, S.112). Solche Strukturierungsversuche laufen demzufolge Gefahr, zu einer theorieleeren Zusammensetzung von empirisch nachweislich wirksamen Faktoren zu verkommen (Einsiedler 1997, Gräsel/Göbel 2015).

4.3.2 Systematik der Prozessmerkmale des Unterrichts

Allein die prominenten Skalen der Unterrichtsforschung zur Erfassung der Unterrichtsqualität aus Schülerperspektive (etwa aus PISA, BIJU u.ä.) variieren erheblich: Sowohl ihre Merkmalsauswahl resp. deren immanenten Qualitätsdimensionen als auch die konkreten Itemformulierungen und -inhalte äquivalenter Merkmale unterscheiden sich voneinander. Demnach wird unter ein- und derselben Rubrik oftmals jeweils unterschiedliches gemessen (Göllner/Wagner/Klieme/Lüdtke/Nagengast/Trautwein 2016).

„Welche Merkmale umfassend und geeignet genug erscheinen, um die Frage nach gutem Unterricht beantworten zu können, ist eine Fragestellung, die im Laufe der letzten Jahrzehnten eine Vielzahl an Kriterienkatalogen und Merkmalslisten zur Folge hatte, die im Vergleich miteinander hohe Überschneidungen aufweisen (Helmke, 2012; Meyer, 2004). [...] Die Kriterienkataloge variieren stark im Hinblick auf die Anzahl [...] und in Bezug auf die Benennung der Kriterien, sodass sowohl weiter als auch enger gefasste Merkmale benannt werden (Williams, 2016) und sich somit Deutungsspielräume eröffnen.“⁶⁹ (Schurig/Gleseemann/Schröder 2016, S. 33)

⁶⁹ Umgang mit Heterogenität und lernförderliches Klima sind demnach beispielhaft für einen ausgeprägten, Klassenführung und Klarheit/Strukturiertheit für einen eingeschränkten Deutungsspielraum (siehe Schurig/Gleseemann/Schröder 2016).

Dennoch lassen sich übergeordnete und fachübergreifend relevante Merkmale, die sich in den unterschiedlichen Kriterienkatalogen widerfinden, individuell je nach Fragestellung ableiten, ohne damit Anspruch auf Allgemeingültigkeit oder Vollständigkeit zu erheben. Die Inhalte der jeweiligen Kriterienkataloge spiegeln dann das jeweilige Verständnis guten Unterrichts wider (Schurig/Glesemann/Schröder 2016). Slavin (1994), Brophy (2000), Meyer (2004) und Helmke (2014) haben die insgesamt einflussreichsten Klassifikationen der Unterrichtsqualität vorgenommen, deren Dimensionalität den meisten Instrumentarien zugrunde liegt. So entspricht guter Unterricht im Allgemeinen einer klaren Strukturierung des Unterrichts, einer klaren Klassenführung, der Schüleraktivierung und -motivierung sowie der individuellen Förderung, einem lernförderlichem Unterrichtsklima und einer Methodenvielfalt zur Differenzierung heterogener Lerngruppen (vgl. Schurig/Glesemann/Schröder 2016, S. 33f). Klieme (2002) unternahm außerdem den Versuch der Komplexitäts- und somit der Dimensionsreduktion entlang wesentlicher Basisdimensionen, die da sind: *kognitive Aktivierung* (v.a. mittels Förderung, Unterstützung, Differenzierung), *Klassenführung* (v.a. durch entsprechendes Setting) und *unterstützendes (Unterrichts-)Klima* (v.a. innerhalb der Lehr-Lernatmosphäre). Alle drei Dimensionen zeigen Effekte hinsichtlich unterrichtlicher Lernziele (wie Leistung, Interesse, Motivation). Allerdings lassen sich identische Daten und Skalen kategorial jeweils unterschiedlich einordnen. Die verschiedenen Modelle erweisen sich dabei entlang von Güteindizes (siehe Kap. 9) als gleichwertig, es ist demnach keines eindeutig einem anderen vorzuziehen (vgl. Willems/Glesemann 2015, Schwanenberg et al. 2015). Ein weiteres Indiz für eine gewisse Willkürlichkeit in der Anzahl und Kategorisierung von Unterrichtsmerkmalen ist die Beobachtung hoher Interkorrelationen zwischen den verschiedenen Merkmalen (siehe Schwanenberg et al. 2015). Bei dieser Feststellung ist jedoch zu betonen, dass sich die dimensionale Struktur der verschiedenen Merkmale dennoch abbilden lässt – sowohl auf Individual- als auch auf klassenebene, was für eine adäquate Differenzierungsfähigkeit der Schülerinnen und Schüler spricht (vgl. Wagner/Göllner/Helmke/Trautwein/Lüdtker 2013). Angesichts der unklaren Befundlage zur Bedeutung einzelner empirisch erfasster Unterrichtsmerkmale für die Unterrichtsqualität wird für diese Arbeit ein Katalog diverser Merkmale verwendet. Die Merkmale sind so ausgewählt, dass sie Befunden verschiedener Verfahren und verschiedenen hinsichtlich der Lernleistungsentwicklung relevanten übergeordneten Merkmalen (wie bspw. *kognitive Aktivierung*, *Klassenführung* und *unterstützendes Klima* nach Klieme, Pauli

und Reusser (in Rakoczy/Buff/Lipowsky 2005)) gerecht werden sowie auf die Eignung von Einschätzungen aus Schülerperspektive validiert sind (vgl. Waldis/ Buff et al. 2002, Lipowsky et al. 2006, Fend/Specht 1986, Clausen 2002, Baumert/ Gruehn/Heyn/Rimmele/Schnabel/Seipp 1996).

4.4 Forschungsüberblick zur Unterrichtsqualität

Im Fokus dieser Arbeit stehen v.a. die Wahrnehmungen der Schülerinnen und Schüler. Danach richtet sich die Beschreibung der Befundlage hinsichtlich der Messung von Unterrichtsqualität entlang von Prozessmerkmalen des Unterrichts aus. Wie oben beschrieben, kann guter Unterricht zwar einerseits nicht allein anhand von Lernergebnissen der Schülerinnen und Schüler ermittelt werden, da solche Lernergebnisse zum einen stark von individuellen Faktoren der Schülerinnen und Schüler abhängen (Jones 1989), zum anderen kann guter Unterricht nicht bedingungslos anhand einer Leistungsnorm ermittelt werden. Andererseits zielt Unterricht vor allem auf die Lernentwicklung der Schülerinnen und Schüler ab. Langzeitstudien zeigen, dass Prozessmerkmale der Unterrichtsqualität durchaus über eine prädiktive Wirkkraft in Bezug auf die Lernentwicklung der Schülerinnen und Schüler verfügen. Dabei ist die genaue Wirksamkeit verschiedener Unterrichtsmerkmale jedoch noch immer unklar und es lassen sich keine einfachen stabilen Abhängigkeitsbeziehungen zwischen isolierten Merkmalen der Unterrichtsqualität und der Lernentwicklung von Schülerinnen und Schülern ausfindig machen. Es ist zu beachten, dass bei derartigen Untersuchungen lediglich der Wahrscheinlichkeitscharakter von Wirkungen im probalistischen und nicht im deterministischen Sinne ermittelt werden kann (vgl. u.a. Ditton/Arnold 2004, Klieme/Rakoczy 2008). Gruehn (2000) konnte diesbezüglich zudem zeigen, dass Schülereinschätzungen der Unterrichtsqualität – meist über eine längere Zeitspanne, auf Basis abstrakter Einschätzungen mit Vermischung von beobachtbaren und innerpsychischen Vorgängen befragt – insgesamt über eine hohe prädiktive Validität verfügen, d.h. einen hohen Erklärungswert für die weitere Lernentwicklung der Schülerinnen und Schüler haben. Lehrereinschätzungen hingegen weisen insgesamt eine geringe prädiktive Validität auf, v.a. da diese Einschätzungen oft stark subjektiv gefärbt und selbstdienlich verzerrt sind. Darüber hinaus ist im Angebot-Nutzungs-Modell Helmkes (s. Abb. 3) die individuelle Wahrnehmung der Schülerinnen und Schüler zwischen Unterrichts-

gebot und dessen Nutzung geschaltet. Demnach wäre von einem Mediationsprozess durch die Schülerwahrnehmungen auszugehen, der sich nicht durch eine von den Schülerperzeptionen losgelöste Erfassung des Angebots eruieren lässt. Um also sowohl der Individualität der Schülerinnen und Schüler und deren Wahrnehmungsperspektive als auch der Bedeutung von Prozessmerkmalen der Unterrichtsqualität für deren Lernentwicklung gerecht zu werden, werden hier vornehmlich Forschungsbefunde, die auf Schülerwahrnehmungen der Unterrichtsqualität basieren, auch sogenannte Schülerfeedbacks, dargestellt. Hierbei ist zu beachten, dass solche Befunde in der Regel durchweg mittels Fragebogenverfahren anhand standardisierter Fragebögen erhoben werden. Dieses Vorgehen ermöglicht es, Kognitionen der Beteiligten innerhalb von Lehr-Lernprozessen zu erfassen, womit es auch eine Annäherung an die Klimaforschung darstellt (Schröder 1999). Darüber hinaus bietet das Fragebogenverfahren Vorteile gegenüber dem (vermeintlich objektiven) Beobachtungsverfahren in Bezug auf die Ausprägung des Erfahrungszeitraums (es ließe sich auch von Beobachtungszeitraum sprechen), die Abbildung heterogener Einschätzungen und einem damit einhergehenden höheren Erklärungswert sowie die ökonomische Handhabung (Morgenstern/Keeves 1993). Zudem können mittels der Erhebung ganzer Klassen und deren aggregierte oder mehrbenenmodellerte⁷⁰ Handhabung als Lernumwelt im Sinne der Klimaforschung Verzerrungen, die vor allem bei Lehrerbefragungen zu uneindeutigen Befunden führen, entgegengewirkt werden (Gruehn 2000, Ditton 2002a). Zwar scheinen die Verzerrungen innerhalb subjektiver Beurteilungstendenzen tendenziell gering, die Wahrnehmung der gesamten Klasse gilt jedoch als in der Tat verlässlich (vgl. auch Stolz 1997). Auf beiden Ebenen – der individuell-spezifischen als auch der kollektiv-geteilten – erweisen sich die Unterrichtsmerkmale als stärkste Einflussvariable auf die Schülereinschätzungen und nicht, wie vormals angenommen, andere subjektive Merkmale. Darüber hinaus wies Clausen (2002) die perspektivenspezifische Validität verschiedener Beurteilungsquellen (Beobachter, Lehrer/Lehrerinnen, Schüler/Schülerinnen)

⁷⁰ Level 1 - subjektive Wahrnehmung des Angebots sowie individuelle Schülermerkmale (bspw. Geschlecht, Jahrgangsstufe, Durchschnittsnoten, sozio-ökonomischer Status, Migrationshintergrund sind als bedeutsame Einflussfaktoren bekannt in der Forschung); Level 2 – wie etwa Schulform, prozentualer Anteil von Schülerinnen und Schülern mit Migrationshintergrund, Schulorganisation (mehr dazu in Kap. 7).

nach. Insgesamt lässt sich aus den signifikanten Unterschieden zwischen Schulklassen ableiten, dass tatsächliche Unterschiede in der Unterrichtsgestaltung bestehen (Ditton 2002a).

Differenzierte Prozessmerkmale des Unterrichts (bspw. Klassenführung, Lernzeitnutzung, Klarheit, Strukturiertheit, Motivierung etc.) korrelieren zudem positiv mit den Lernerfolgen der Schülerinnen und Schüler (Helmke 2003, Lipowsky 2006). Ein bestimmtes Maß an fachdidaktischen Kompetenzen der Lehrkraft stellt allerdings den stärksten Prädiktor für einen auf Klassenebene aggregierten Lernzuwachs dar (vgl. Hill/Rowan/Ball 2005, Baumert 2006 (COACTIV)). Außerdem zeigt sich ein Lernzuwachs von Schülerinnen und Schülern nach bestimmten Lehrerfortbildungen (z.B. das Münsteraner BIQUA-Projekt, vgl. Kleickmann/Müller 2007), wenn fachdidaktisches, inhaltsbezogenes, curriculares Lehrerwissen durch die Fortbildung erweitert wurde, was auf die Relevanz von fachdidaktischen resp. inhaltlichen Aspekten des Unterrichts (Prozessmerkmale der *kognitiven Strukturierung* und *Aktivierung*) für den Lernzuwachs hinweist (siehe hierzu auch Lipowsky 2006, Einsiedler/Treinies 1997, Shayer/Adhami 2007).

4.4.1 Schülerperzipierte Unterrichtsqualität

Obwohl bereits 1915 eine erste Skala zur Beurteilung von Lehrerinnen und Lehrern aus Schülersicht veröffentlicht worden war, nahmen die Untersuchungen zur Schülerwahrnehmung des Unterrichts sowie deren Einflussvariablen erst 1920 mit Remmers ihren Anfang. Zunächst wurden Verfahren der Schülerevaluation von Unterricht in US-Universitäten vereinzelt, dann zunehmend eingeführt (Wachtel 1998). Seit der empirischen Wende erfuhr die Unterrichtsforschung ab den 1970er Jahren neuen Antrieb durch Studien zur Validität und Nützlichkeit von Klassenklimaevaluationen aus Schülersicht, die den vermehrten Einsatz von Schülerperspektiven zur Unterrichtsevaluation⁷¹ unterstützten. Klassenklimaevaluationen implizieren zwar auch Unterrichtsmerkmale, beziehen sich ursprünglich jedoch vornehmlich auf die dem Lernklima zugrunde liegenden Schüler-Schüler- und Lehrer-Schüler-Beziehungen. Zudem finden sich

⁷¹ Drei Validierungsansätze der Evaluationsforschung (Greimel 2002):

1. Per Lehrziele (=Lernerfolge)
2. Konstruktvalidierung von Unterrichtsqualität mittels Einbindung u.a. affektiver Unterrichtsziele (geht über den Ansatz des Lernerfolgs hinaus)
3. Per Biasfaktoren.

Untersuchungen in der Forschungstradition der Klimaforschung v.a. in der Hochschulforschung im amerikanischen Raum, die sich erst innerhalb der letzten zwei Jahrzehnte in Richtung Unterrichtsevaluation aus Schülersicht entwickelte. Im Laufe der 1980er wurden die Forschungsergebnisse der verschiedenen Forschungstraditionen (hochschulische Evaluationsforschung, empirische Unterrichtsforschung, Klassenklimaforschung) noch detaillierter geklärt und die Qualität von Forschung sowie deren Veröffentlichungen gesteigert. Damit einhergehend sowie mit der zunehmenden Konvergenz der verschiedenen Forschungstraditionen hinsichtlich der Berücksichtigung der Schülerperspektive wurde die Forschung in Bezug auf Schülerwahrnehmungen weiter gestärkt (Wachtel 1998, Marsh/Dunkin 1992). Aleamoni stellte 1981 heraus, dass die Meinungen zu Schülerurteilen zwar auseinanderklaffen und von reliabel, valide und nützlich bis hin zu unreliabel, invalide und nutzlos reichen. Doch schätzten die meisten Forscher Schülerurteile positiv ein. Entsprechend konstatierte Marsh bereits 1984, dass die Validität von Schülerurteilen ausreichend bewiesen wurde, sodass der Forschungsfokus zunehmend andere Aspekte, die diese Urteile beeinflussen, in den Blick nehmen könne. Validität und Reliabilität resp. Korrektheit von Schülerbeurteilungen des Unterrichts sind jedoch auch weiterhin umstritten (nach Wagenaar 1995) und so finden sich in Diskussionen weiterhin sowohl für als auch wider die Validität von Schülerbeurteilungen Argumente (Auswahl aus Wachtel 1998, S. 192ff und Göllner/Wagner/Klieme/Lüdtke/Nagengast/Trautwein 2016, S. 66):

Pro-Argumente besagen bspw., dass

- Schülerbeurteilungen differenziert erfolgen sowie effektiv und flexibel einsetzbar sind (Göllner et al. 2016).
- substantielle Zusammenhänge zwischen Schülerbeurteilungen des Unterrichts und Kriterien des Schulerfolgs bestehen (Gruehn 2000, Clausen 2002, Klieme et al. 2008), wonach Schüler und Schülerinnen am meisten bei jenen Lehrerinnen und Lehrern lernen, die sie positiv einschätzen und beurteilen (Aleamoni 1981, Cohen 1981, Centra 1977, Murray et al. 1990).
- Schülerfeedbacks helfen, Unterricht zu verbessern⁷² (Cohen 1980, Marsh et al. 1993).

⁷² Kritik an diesem Punkt wird mit dem Argument der diesbezüglich schwachen Forschung begegnet (L'Hommedieu et al. 1990).

- der Einsatz von Schülerfeedbacks die Wahrscheinlichkeit erhöht, gute Lehre zu erkennen und zu honorieren (Aleamoni 1981).
- Schülerinnen und Schüler mit ihren Lehrkräften grundlegend in ihren Kriterien bezüglich guten (resp. effektiven) Unterrichts sowie deren Bedeutsamkeit übereinstimmen (Feldmann 1976 & 1988, Marsh & Dunkin 1992).
- Schülerurteile positiv mit denen Ehemaliger korrelieren und demnach valide sind (vgl. bspw. Feldmann 1989, Centra 1974).

Contra-Argumente hingegen beziehen sich z.B. darauf, dass:

- keine adäquate Evaluation von Unterricht möglich ist, da keine einheitliche Definition darüber existiert, was effektives Lehren ausmacht (resp. was guten Unterricht bedingt) (vgl. Marques et al. 1979, Spencer 1992).
- Evaluationen Kapazitäten, die für innovativen Unterricht genutzt werden könnten, beschneiden (Marques et al. 1979, Spencer 1992).
- keine Fächervergleiche von Unterrichtsevaluationen aus Schülersicht möglich sind (Göllner et al. 2016).
- Schülerinnen und Schüler nicht über eine ausreichende Differenzierungskompetenz verfügen, um Unterricht professionell beurteilen zu können (vgl. Göllner et al. 2016).
- Schülerurteile nicht objektiv sind, da die beurteilenden Schülerinnen und Schüler selbst in das Unterrichtsgeschehen involviert sind (vgl. Göllner et al. 2016).
- die meisten Evaluationsinstrumente unsachgemäße Items (vage, mehrdeutig, subjektiv gefärbt) beinhalten oder solche, die nicht im Zusammenhang mit der Unterrichtsqualität stehen (Marques et al. 1979, Spencer 1992, Göllner et al. 2016).
- Feedbacks generell durch Biasvariablen verzerrt sind (Marques et al. 1979, Spencer 1992).

Es lässt sich festhalten, dass es zur Entscheidung zwischen den Argumentationslinien noch immer in weiten Teilen an Forschungsbefunden fehlt. So zeigen Studien zwar z.B., dass Hintergrundvariablen Ergebnisse von Lehrereinschätzungen durchaus signifikant verändern können. Allerdings bedeutet nicht jede Korrelation mit einer oder mehreren Hintergrundvariablen direkt eine Verzerrung oder verminderte Validität der Aussagen. Dies wird am Beispiel von Zensuren anschaulich: So bedingen schlechte Noten in der Regel zwar durchaus eine schlechte Evaluation, gute Noten bedingen hingegen nicht gleichermaßen eine gute Evaluation. An dieser Stelle zeigt sich der Bedarf an präziser Forschung zu

Biasvariablen von Schülerfeedbacks (Brandenburg et al. 1977, Tatro 1995). Einige grundlegende Befunde lassen sich jedoch unabhängig davon, entlang welcher Prozessmerkmale des Unterrichts genau Schülerwahrnehmungen eingeholt werden, nachweisen. Manche sind äußerst naheliegend, wie etwa die Feststellung, dass sich schülerspezifische Wahrnehmungen am besten anhand von Erhebungen der Schülerwahrnehmungen abbilden lassen (Ditton 2002a). Andere sind erfreulich wie der Befund, dass insgesamt eine positive Wahrnehmung des Unterrichts (insbesondere der Variablen *Interessantheit*, *Klarheit*, *Diagnose* und *Schüler-Lehrer-Verhältnis*) durch die Schülerinnen und Schüler zu verzeichnen ist (Ditton/Arnold 2004, Seifried 2006) oder unerfreulich wie der weitgehend gesicherte Befund hinsichtlich der Variationsarmut eingesetzter Unterrichtsmethoden (Ditton 2002). Darüber hinaus ist festzuhalten, dass sich hinsichtlich der methodischen Unterrichtsgestaltung⁷³ (Seifried 2006) sowie der Klassenführung im Sinne von Störungsfreiheit und Zeitnutzung (Kunter/Baumert 2006, Gruehn/Bertram 2014) hohe Übereinstimmungen zwischen den Wahrnehmungen der Lehrenden und der Lernenden nachweisen lassen. Insgesamt überwiegen jedoch hinsichtlich diverser Unterrichtsmerkmale, verschiedener Fächer und Lerninhalte Unterschiede in den Differenzen zwischen den Schüler- und Lehrerwahrnehmungen (Gruehn 2000, Wagner 2010, Seifried 2006, Klieme 2006, Wagner 2008, Baumert/Kunter 2003, 2006).

Ditton (2002a) zeigte zudem, dass je jünger die befragten Schülerinnen und Schüler sind, desto undifferenzierter ihre Angaben zur Unterrichtsqualität erfolgen. Das heißt, dass jüngere Schülerinnen und Schüler nicht so ausgeprägt wie jene älterer Jahrgänge zwischen verschiedenen Merkmalen oder Dimensionen des Unterrichts unterscheiden. Neuere Befunde (Schurig et al. 2016) weisen jedoch darauf hin, dass bei einer adäquaten Erfassung von Unterrichtsqualität anhand von strukturierten Prozessmerkmalen Schülerurteile durchaus vergleichbar über verschiedene Altersstufen hinweg eingesetzt werden können. Allerdings nehmen Schülerinnen und Schüler älterer Jahrgänge (in dem Fall der Schulstufe acht) gegenüber jüngeren (Schulstufe drei) den Unterricht insgesamt als weniger interessant und positiv sowie ihren Entscheidungsspielraum als kleiner wahr (Gentry/Rizza/Gable 2002). Entlang der drei Basisdimensionen *Klassenführung*

⁷³ Hierbei ist zu beachten, dass die Methodenwahl zwischen den Fächern durchaus variiert (vgl. Seifried 2006).

im Sinne strukturierter Unterrichtsführung, *Schülerorientierung* und *kognitiver Aktivierung* (denen zahlreiche etablierte Instrumente zugrunde liegen) zeigt sich zudem, dass die Dimension *Schülerorientierung* (als Prozessmerkmal der Unterrichtsqualität) in höheren Jahrgängen schlechter bewertet wird als von jüngeren Schülerinnen und Schülern.⁷⁴ Über alle Jahrgänge scheint jedoch eine gute Schüler-Lehrer-Beziehung mit guten Lehr-Lern-Prozessen einherzugehen (Radisch/Stecher/Klieme/Kühn-bach 2007).

In Bezug auf Geschlechtsunterschiede zeigen die Befunde, dass Mädchen über die Schulstufen drei bis acht hinweg den Unterricht offenbar zwar durchweg als interessanter und positiver empfinden als es die Jungen kundtun (Gentry/Rizza/Gable 2002). In anderen Untersuchungen, die auf Mehrebenenmodellen beruhen, korrelieren wesentliche individuelle Merkmale wie das Geschlecht, der sozio-ökonomische Hintergrund, die Fachnote und die Einstellung zur Schule jedoch kaum mit Einschätzungen des Unterrichts.⁷⁵ Ebenso verlieren strukturelle Merkmale wie die Klassengröße⁷⁶ an Gewicht und weisen keine besonderen Signifikanzen mehr auf (Ditton 2002a). Insgesamt zeigt sich hinsichtlich des Einflusses des Geschlechts eine recht uneinheitliche Forschungslage.

In Bezug auf die Klassenmerkmale wird berichtet, dass Wahlkurse besser als Pflichtkurse bewertet werden. Wenn zudem die Mehrheit der Klasse den Kurs als *freiwillig* (möglicherweise durch ein erhöhtes Interesse) erlebt, verstärken sich die Evaluationswerte den Wahlkursen entsprechend (Feldmann 1978, Brandenburg 1977). Unklare Befunde bestehen bzgl. möglicher positiver Korrelation zwischen Arbeitspensum und Unterrichtsqualität sowie zum direkten oder indirekten Einfluss der Klassengröße.

Zudem fallen Bewertungen besser aus, wenn die Feedbacks persönlich zuzuordnen sind; schlechter hingegen, wenn sie anonym erfolgen (Feldman 1979, Blunt 1991). Auch die Anwesenheit des Lehrers bzw. der Lehrerin während der Erhebung verbessert die Feedbackwerte (Feldmann 1979), wobei ein Zusammenhang

⁷⁴ Das wird dahingehend interpretiert, dass Angebote möglicherweise die Interessen-/Bedürfnislage der Schülerinnen und Schüler verfehlen (Radisch/Stecher/Klieme/Kühnbach 2007).

⁷⁵ Dieser Befund bezieht sich auf eine Studie, die Unterricht auf dem von Slavin (1994) entwickelten QAIT-Modell erhebt, das sich auf die Dimensionen Qualität (quality), Angemessenheit (adaptivity), Anregung (interesting), Zeitnutzung (use of time) bezieht.

⁷⁶ Zuvor wurde nachgewiesen, dass auf Aggregatebene kleinere Klassen dazu tendieren, den Unterricht besser einzuschätzen als größere Klassen (Ditton 2002a).

mit der Anonymität vermutet wird (Centra 1993). Daher sollten Dritte der Erhebung beiwohnen und die Fragebögen einsammeln.

Um sich innerhalb der recht unsystematischen Verfahren zur Erhebung von Schülereinschätzungen zur Unterrichtsqualität besser orientieren und deren Reliabilität und Validität besser einordnen zu können, prüfte die Forschungsgruppe (Göllner/Wagner/Klieme/Lüdtke/Nagengast/Trautwein 2016) der Tübinger Universität (Hector-Institut für Empirische Bildungsforschung) Möglichkeiten der angemessenen Erfassung der Unterrichtsqualität aus Schülersicht. Dabei sind die den Items immanenten Auslegungen der Adressaten, der Perspektive und des zeitlichen Rahmens von besonderem Interesse: die Adressaten und Perspektiven sind in den gängigen Skalen bislang eher mehrdeutig gestaltet und lassen Interpretationsspielraum für die Befragten offen; ebenso ist dies hinsichtlich des zeitlichen Rahmens oftmals der Fall – mit dem Effekt, dass durchaus anzunehmende Veränderungen im Lehrerverhalten während des Schuljahres zu variierenden Beurteilungen durch die Schülerinnen und Schüler führen. Der zeitliche Rahmen sollte demzufolge in die Fragestellungen einbezogen werden, um die Varianz der Schülereinschätzungen zu verringern. Es wird zwar außerdem empfohlen Unterrichtsbeurteilungen mehrerer Messzeitpunkte zusammenzufassen, da sich kein Messzeitpunkt als optimal auszeichnet. Allerdings variiert der Einfluss des Messzeitpunkts auf die Aussagekraft von Zielvariablen je nach gemessenem Unterrichtsmerkmal unsystematisch – so ist bspw. die Zeitspezifität für das Merkmal Klassenführung besonders ausgeprägt, während andere Merkmale nur gering beeinflusst sind (Göllner/Wagner/Klieme/Lüdtke/Nagengast/Trautwein 2016). Frühere Untersuchungen diesbezüglich zeigten in Bezug auf die Durchführung der Erhebung selbst kaum Unterschiede in den Ergebnissen (Feldmann 1979).

Darüber hinaus stellen bei der Itemformulierung sprachliche und kontextuelle Merkmale mögliche Ursachen für die Varianz zwischen Schülereinschätzungen dar. So zeigt sich in der Untersuchung von Göllner, Wagner et al. (2016), dass sowohl die Anzahl von Nebensätzen und Nominalisierungen die Abstände zwischen den Schülerurteilen vergrößert. Daraus zieht die Forschergruppe den Schluss, dass die Schülerinnen und Schüler möglicherweise dieselben Meinungen zur Unterrichtsqualität haben, jedoch die Fragestellungen unterschiedliche interpretieren. Auf andere sprachliche Merkmale wie die Zeichenanzahl pro Wort oder Wortanzahl pro Satz ließen sich keine entsprechenden Rückschlüsse ziehen. Insgesamt zeigt sich in der Untersuchung, dass Schülereinschätzungen

der Unterrichtsqualität stärker mit unterrichtlichen Zielvariablen zusammenhängen als Lehrer- oder Beobachterdaten, was die Bedeutsamkeit von Schülerfeedbacks verstärkt (Göllner/Wagner/Klieme/Lüdtke/Nagengast/Trautwein 2016).

4.4.2 Fachspezifität

Noch immer stellt sich bei Schülerbefragungen zur Unterrichtsqualität neben der Frage nach deren Dimensionalität die Frage nach deren Fachspezifität bzw. -unabhängigkeit. Um dahingehend Einblicke zu gewinnen, wurden Schülerdaten (der Jahrgänge fünf, sieben und neun) des Projekts *Ganz In* (siehe hierzu bspw. Berkemeyer/Bos/Holtappels/Meetz/Rollett 2010) auf Individualebene⁷⁷ ausgewertet. Dabei wurden sieben Unterrichtsmerkmale in den Fächern Biologie, Deutsch, Mathematik und Englisch berücksichtigt. Zwischen den Messzeitpunkten sowie zwischen den Fächern konnte eine Messinvarianz abgeleitet werden, demzufolge in verschiedenen Fächern und Altersstufen den latenten Variablen identische inhaltliche Bedeutungen zugeschrieben werden. „Damit ist die Übertragbarkeit von einem Fach auf ein anderes Fach grundsätzlich gegeben, wie von Helmke (2012) angenommen wurde“ (Schurig/Glesemann/Schröder 2016, S. 49). Allerdings bedeutet dies nicht, dass Items in unterschiedlichen Fächern identisch interpretiert und somit gleichermaßen beantwortet werden – eine unerfüllte Voraussetzung für einen adäquaten Fächervergleich.

Diverse Untersuchungen zeigen, dass Unterrichtsqualität fächergetrennt betrachtet werden muss bzw. von den Schülerinnen und Schülern je nach Fach unterschiedlich eingeschätzt wird (bspw. Gruehn 2000, Klieme 2006, Wagner et al. 2013, Rahn et al. 2016). Das heißt, dass Aspekte der Prozessqualität des Unterrichts zwar fachübergreifend gewisse Grunddimensionen enthalten, diese jedoch je nach Fach spezifisch und unabhängig voneinander ausgelegt werden (vgl. auch Klieme/Rakoczy 2008):

„Eine Verschlechterung der von den Schülerinnen und Schülern wahrgenommenen Unterrichtsqualität ist für die Fächer Englisch und Mathematik festzustellen, nicht dagegen für das Fach Deutsch. [...] Insgesamt fallen die

⁷⁷ Im Rahmen dieser Untersuchung wurde mit Mplus anhand des Analysebefehls *complex* gearbeitet, der die hierarchische Struktur der Daten insofern berücksichtigt, als dass Standardfehler und damit die Modelgüteindizes auf Individualebene entsprechend korrigiert werden.

Ergebnisse für das Fach Mathematik ungünstiger aus als für die Lehrkräfte der beiden anderen im Längsschnitt untersuchten Fächer.“ (Ditton/Arnold 2004, S. 159)

Unterrichtsfächer wie Mathematik und Naturwissenschaften verzeichnen stets tendenziell niedrige Ratingwerte (Ramsden 1991), woraus sich ergibt, dass daher keine fachübergreifenden Vergleiche möglich sind. Allerdings lassen die Befunde nicht ausschließlich den Rückschluss zu, dass nur die Schülerinnen und Schüler es sind, die den Unterricht fachspezifisch einschätzen. Mit der DESI-Studie wurde gezeigt, dass die zwar fachübergreifend bedeutsamen Merkmale je nach Fach didaktisch unterschiedlicher Gestaltung bedürfen: Bspw. müssen Aspekte der kognitiven Aktivierung in sprachlichen Fächern anders gestaltet werden als im Fach Mathematik (Klieme/Rakoczy 2008).

In PISA 2006 wurden außerdem Übereinstimmungen von Schülerinnen und Schülern innerhalb ihres Klassenverbandes in Bezug auf den naturwissenschaftlichen Unterricht mit dem Ergebnis überprüft, dass diese ihren Unterricht innerhalb der Klasse zwar konsistent und reliabel beurteilen, die klasseninternen Übereinstimmungen jedoch je nach Klassengröße, effektiver Unterrichtszeit, Schulart und vor allem entsprechender Fachkompetenz der Schülerinnen und Schüler (gemessen am Klassendurchschnitt) variieren. Wittwer (2008) fordert daher von der Forschung, die Wahrnehmungsunterschiede innerhalb von Klassen verstärkt zu berücksichtigen. Diesbezüglich weisen neuere Forschungen (Wagner et al. 2013) darauf hin, dass verschiedene Dimensionen⁷⁸ unterschiedlich auf verschiedenen Ebenen funktionieren: Dimensionen mit hohem Schülerbezug wie *Schülerorientierung*, *Motivierung* oder *Verständlichkeit* fallen deutlich lehrkraftspezifischer aus als die Dimensionen *Strukturiertheit* und *Klassenführung*. Zumal Forschungen außerdem die fachspezifische Relevanz von Unterrichtsmerkmalen für entsprechende Lernerfolge aufzeigen (Seidel/Shavelson 2007). Die Bedeutung von Merkmalen des Unterrichts variiert demnach ebenso je nach Fach wie deren Vorkommen und deren Wahrnehmung.

⁷⁸ Die Untersuchung bezieht sich auf Daten der DESI-Studie und verwendet aufgrund der Komplexität der Modellierung lediglich die Unterrichtsmerkmale Schülerorientierung, Motivierung, Verständlichkeit, Strukturiertheit und Klassenführung (siehe Wagner et al. 2013).

5 Ableitung von Forschungsdesiderata und Fragestellungen

Persönlichkeitseinschätzungen hängen davon ab, aus welcher Perspektive sie erhoben werden, wobei keiner Perspektive Anspruch auf universelle Gültigkeit zugesprochen werden kann. Innerhalb der Persönlichkeitsforschung resp. der Differentiellen Psychologie bietet sich mit der Systematik der BigFive ein theoretisch hinreichend begründetes Konstrukt, mit dem die Persönlichkeit umfassend auf Basis von Eigenschaften bzw. Dispositionen sowohl selbst- als auch fremdeingeschätzt erfasst werden kann. In Assessment Centers findet der Einsatz des BigFive-Konstrukts zur persönlichen Selbsteinschätzung in Verbindung mit der Tauglichkeit für den Lehrerberuf (wie bereits für andere Berufe üblich) vermehrt Einzug. Allerdings zielt hierbei der Fokus v.a. auf die Frage nach der Belastbarkeit und der Zufriedenheit im Beruf ab – nicht direkt auf die Unterrichtsqualität, die jedoch als eine selbstverständliche Folge dessen angesehen wird. Untersuchungen zur Bedeutung persönlicher Merkmale beruhen generell fast ausschließlich auf Selbsteinschätzungen. Die Unterrichtsforschung bildet bei der Erfassung der Lehrerpersönlichkeit diesbezüglich keine Ausnahme. Die systematische Überprüfung des Zusammenhangs von selbsteingeschätzter Lehrerpersönlichkeit und jener aus Schülerperspektive bleibt demnach ein Forschungsdesiderat. Folge dieser Forschungslücke ist eine perspektivenübergreifende Vermischung von Wahrnehmungen, wenn die Lehrerpersönlichkeit mit der Unterrichtsqualität in Verbindung gebracht wird. Die Unterrichtsqualität wird zunehmend in Form von Unterrichtsevaluationen aus Schülersicht erfasst, während die Lehrerpersönlichkeit in der Regel aus Lehrerperspektive erfragt wird. In Forschungen zur Erfassung von *Persönlichkeit*, die den Zusammenhang unterschiedlicher Perspektiven untersuchen, zeigen sich jedoch überwiegend schwache bis moderate Zusammenhänge zwischen selbst- und fremdeingeschätzten Persönlichkeitsmerkmalen (entlang der BigFive-Systematik).⁷⁹ Insgesamt zeigen sich zudem merk-

⁷⁹ Es finden sich Befunde, die von mittleren Korrelationen von 0,30-0,60 (vgl. Fremdbbericht in Wirtz 2014, Vazire/Carlson 2010) berichten sowie solche, die von schwachen Korrelationen von 0,03-0,37 berichten (vgl. Watson/Hubbert/Wiese 2000, Feldmann 1986).

malsspezifische und je nach Grad der Bekanntheit Unterschiede in der Ausprägung des Zusammenhangs verschiedener Perspektiven.⁸⁰ Es lässt sich lediglich eine Studie finden, die diesen Zusammenhang zwischen Lehrer- und Schülerperspektiven untersucht (vgl. Mayr/Neuweg 2006, außerdem siehe Kap. 2.2.3), dies allerdings ohne einen Fachbezug.

Zwar gibt es vereinzelte Studien, die sich auf die Schülerperspektive bei der Verknüpfung von Lehrerpersönlichkeit und Unterrichtsqualität konzentrieren, allerdings erfassen diese Erhebungen vor allem amerikanische Universitäten, deren Unterricht kaum mit dem schulischen Unterricht in Deutschland vergleichbar ist. Erschwerend für die Ableitung von Schlussfolgerungen, die dem Standard empirischer Forschung genügen, kommt hinzu, dass die Stichproben besagter Studien durchweg sehr klein sind und/oder das Forschungsdesign durchweg unsystematisch erscheint. Die Metaanalyse von Feldmann von 1986 bildet insofern eine Ausnahme, als bis dato bestehende Untersuchungen systematisch aufbereitet, verglichen und zusammengefasst bewertet werden. Dabei zeigt sich, dass diverse Fremdeinschätzungen der Lehrerpersönlichkeit (d.h. aus Schüler- sowie Kollegenperspektive) mit schülerperzipierten Unterrichtseffektivitätseinschätzungen bei den meisten der 14 erhobenen Persönlichkeitsmerkmale durchaus moderat bis stark (,50 und höher) korrelieren. Feldmann schließt aus diesem Befund seiner Metaanalyse, dass es in Bezug auf den Zusammenhang von Lehrerpersönlichkeitsmerkmalen mit der Unterrichtsqualität angebracht scheint, verschiedene Perspektiven auf die Lehrerpersönlichkeit getrennt voneinander zu analysieren. Zumal Feldmann den Wahrnehmungen der Schülerinnen und Schüler zwar Validität zugesteht, allerdings unter der – auch in dieser Arbeit voraussetzungsvollen – Prämisse, dass es sich dabei nicht um die Beschreibung einer *wahren* Persönlichkeit handelt, sondern lediglich um eine spezifische und kontextuelle Perspektive⁸¹. Darüber hinaus zeigen seine Analysen, dass Schülerinnen und Schüler zwischen Unterrichtseffektivität und Lehrerpersönlichkeit gut

⁸⁰ Bspw. zeigen sich beim Persönlichkeitsmerkmal *Neurotizismus* unter weniger guten Bekanntschaften Korrelationen von 0,27 auf, bei besseren Bekanntschaften von 0,46 (vgl. Asendorpf/Neyer 2012); beim Merkmal *Gewissenhaftigkeit* zeigen enge Freundschaften Korrelationen von 0,39 und bei *Extraversion* von 0,54; die Werte bei den Persönlichkeitsmerkmalen *Offenheit* und *Verträglichkeit* korrelieren zwischen 0,40 und 0,49 (Back/Vazire 2012)).

⁸¹ Durch die Aggregation auf Klassen- bzw. Lehrkräfteebene weist diese Perspektive allerdings sowohl eine höhere Reliabilität als auch geringere Verzerrung auf (vgl. Feldmann 1986).

zu differenzieren verstehen und keinesfalls das eine mit dem anderen vermischen. Allerdings besteht die Möglichkeit, dass Persönlichkeitseinschätzungen ebenso wie Effektivitätseinschätzungen durch die Schülerinnen und Schüler denselben Biasvariablen unterliegen und dadurch die Höhe der Korrelationen zustande kommt (Feldmann 1986). Bis dato finden sich kaum entsprechende systematische perspektivenspezifische Forschungen zum Zusammenhang der Lehrerpersönlichkeit und der Unterrichtsqualität, die beide Konstrukte zum einen systematisch und differenziert und zum anderen aus Schülerperspektive erfassen.

Weitere Untersuchungen weisen nach, dass die mittels gegenseitigen Peer-Ratings erfasste Lehrerpersönlichkeit einen signifikanten Prädiktor für effektiven Unterricht bzw. für positive Schülerratings bzgl. der Lehreffektivität darstellen kann (Schmidt et al. 1991, Murray et al. 1990, Magno/Sembrano 2008). Renaud und Murray (1996) zeigten diesbezüglich, dass die Lehreffizienz zwar signifikant mit einzelnen Persönlichkeitseigenschaften korreliert, jedoch nicht mit *Extraversion* oder *Neurotizismus*. Eine weitere Studie (von Clayson/Sheffet 2006, Details siehe Kap. 2.2) stellt heraus, dass Persönlichkeitsdimensionen – wiederum außer *Extraversion* – signifikante Regressionskoeffizienten in Bezug auf einen übergeordneten Faktor *Evaluation* aufweisen, sie also starke Prädiktoren sind. Ebenso zeigt die Studie, dass die *Unterrichtsevaluation* mit Veränderungen der Wahrnehmung der Lehrerpersönlichkeit variiert, dies allerdings v.a. hinsichtlich der Wahrnehmung von *Extraversion*. Korrelationen zwischen dem übergeordneten Persönlichkeitsfaktor und der erhobenen Unterrichtsevaluation nehmen zudem über vier Erhebungszeitpunkte immer stärker zu, wobei sie durchweg signifikant sind. Im Fach Deutsch können diese Befunde hinsichtlich der BigFive-Dimension *Extraversion* bestätigt werden: lediglich auf das Prozessmerkmal *Störungsprävention* ist ein zwar signifikanter, jedoch moderater Effekt (0,29) zu verzeichnen. Allerdings ist dieser Befund keineswegs fachübergreifend: Im Fach Mathematik zeigt die Dimension *Resilienz* die wenigsten (drei im Gegensatz zu vier bzw. sechs moderat bis starken Effekten der anderen vier Persönlichkeitsdimensionen) und insgesamt für das Fach Mathematik betrachtet auch die schwächsten Effekte (mit zwei knapp moderaten Regressionskoeffizienten um 0,30 – auf die *Methodenvariation* und die *Motivierungsqualität* – und einem deutlicheren Effekt von 0,52 auf die *Disziplin*) auf die Prozessmerkmale des Unterrichts. *Extraversion* stellt im Fach Mathematik jedoch ebenso wie die *Resili-*

enz im Fach Deutsch deutliche Prädiktoren (0,26-0,67) für vier von sieben Prozessmerkmalen des Unterrichts. Allerdings werden auch diese Studien dem Desiderat der Perspektivenspezifität nicht gerecht, sondern stellen Zusammenhänge perspektivenübergreifend her.

Vor dem Hintergrund der Studie von Klusmann et al. (2006, siehe hierzu den entsprechenden Abschnitt in Kap. 2.2), die im Fach Mathematik anhand von selbsteingeschätzten Lehrerbelastungstypen nach AVEM deutliche Effekte des positiven und als resilient zu bezeichnenden Gesundheitstyps auf vier von sechs Prozessmerkmalen der von den Schülerinnen und Schüler wahrgenommenen Unterrichtsqualität ausmachen konnten, lassen sich außerdem die hier beobachteten Effekte der *Resilienz* (ebenfalls in Mathematik) interpretieren. Zum einen dahingehend, dass die Wirkung der *Resilienz* der Lehrkräfte auf die schülerperzipierte Unterrichtsqualität trotz differenzierter Erhebungsformen Ähnlichkeiten aufweisen: sowohl die im Lehrerselbstbericht erfassten AVEM-Typen als auch die durch die Schülereinschätzungen erfasste Lehrerpersönlichkeit entlang der BigFive zeigen keine Effekte der *Resilienz* der Lehrkräfte auf solche unterrichtlichen Aspekte, die sich auf Störungen im Unterrichts beziehen (*Störungen im Unterricht* (Klusmann et al. 2006) resp. *Störungsprävention*), jedoch durchaus auf andere Unterrichtsaspekte. Zum anderen lässt sich möglicherweise die Big-Five-Dimension *Resilienz* vergleichbar oder ggf. noch differenzierter anhand der AVEM-Typen erfassen. Da aus Schülersicht die Unterrichtsqualität von ein und derselben Lehrkraft fachübergreifend variieren kann und zudem gewisse persönliche Merkmale offenbar in einem Fach bedeutsam für die Unterrichtsqualität und im anderen bedeutungslos sein können (vgl. auch Murray et al. 1990), lassen sich die Ergebnisse von Klusmann et al. (2006) jedoch nicht auf andere Fächer übertragen. Daher wird sich an dieser Stelle der Arbeit an den Schluss von Norzila et al. (2007) angelehnt, demzufolge es auf die Passung zwischen Unterrichtsstil und Persönlichkeit ankommt, um effektiv zu lehren. Dabei ist v.a. zu beachten, dass sich nicht jeder Stil gleichermaßen für jedes Fach anbietet. Es bleibt demnach eine offene Frage, wie sich derartige Zusammenhänge fachspezifisch auswirken. Diese Arbeit verknüpft daher die Prozessorientierung der Lehr-Lernforschung und die dispositionale Persönlichkeitsforschung, die bislang v.a. in Hinblick auf Assessment Centers angewandt wird, innerhalb der Unterrichtsforschung. Dabei werden hinsichtlich der Persönlichkeitserfassung (anhand der BigFive) etablierte Ansätze aus der differentiellen resp. Persönlichkeitspsychologie entliehen. Diese werden erstmals systematisch auf die perspektivenspezifische Wahrnehmung der

Lehrerpersönlichkeit (durch die Schülerinnen und Schüler und deren Lehrkräfte) sowie auf die Unterrichtsforschung zur Unterrichtsqualität aus Schülersicht übertragen. Da das Instrument zur Ermittlung der BigFive (BFI-S) erstmals zur Fremdeinschätzung und innerhalb der Unterrichtsforschung eingesetzt wird, beschreibt dessen Validierung einen Vorbefund, der für die weiteren Fragestellungen der Arbeit Voraussetzung ist.

Weitere grundlegende Vorbefunde finden sich daran anschließend in den deskriptiven Analysen der erhobenen Konstrukte, die sich an einleitenden Fragestellungen orientieren: Wie schätzen sich die Lehrkräfte persönlich ein? Wie nehmen die Schülerinnen und Schüler der elften Jahrgangsstufe die Lehrerpersönlichkeit und wie den Unterricht in den Fächern Deutsch und Mathematik wahr? Bestehen in den Wahrnehmungen der Konstrukte über fachspezifische Besonderheiten hinaus auch geschlechtsspezifische Tendenzen?

Zunächst gilt es jedoch angesichts des innerhalb der Schulforschung vorgefundenen doppelten Forschungsstandes zur Lehrerpersönlichkeit – d.h. einer Diskrepanz zwischen aktuellem theoretischen und entsprechend empirischem Forschungsstand – die Selbst- und Fremdwahrnehmung (durch die Schülerinnen und Schüler resp. die Klasse) der Persönlichkeit von Lehrkräften empirisch zu prüfen, um dadurch eine perspektivenspezifische Persönlichkeitswahrnehmung aufzudecken. Damit wären solche Befunde wie jene von Mayr und Neuweg (2006), die sich derzeit innerhalb von Assessment Centers etablieren, insofern zu relativieren, als diese keine direkten Rückschlüsse auf die Qualität des Unterrichts zuließen. Es stellt sich in einem ersten Schritt also die Frage nach dem korrelativen Zusammenhang von Lehrerselbsteinschätzung und Schülereinschätzungen der Lehrerpersönlichkeit. Demnach lautet die erste Forschungsfrage dieser Arbeit:

Forschungsfrage 1: Welcher Zusammenhang besteht zwischen selbst- und fremdwahrgenommener Lehrerpersönlichkeit?

Ein wesentlicher Nachteil des sehr variablenzentrierten Vorgehens zur Ermittlung der Unterrichtsqualität anhand von Prozessmerkmalen besteht darin, dass persönliche Merkmale der Lehrenden dabei weitgehend ausgeblendet werden. Gleichzeitig zeichnet sich ein neu entflammter alter Trend innerhalb der Unterrichtsforschung als auch der Lehrerausbildung ab, der die Lehrer und Lehrerinnen wieder vermehrt als Person in den Blick nimmt und diese als ausschlaggebend für guten Unterricht identifiziert. Allerdings bewegen sich solche auf die Lehrerpersönlichkeit gerichteten Aussagen nahezu alle in der Nähe von Behauptungen: a) gründen sie entweder auf selbstberichteten Lehrerpersönlichkeiten oder

aber b) auf untersuchten Merkmalen, die im etablierten Persönlichkeitsverständnis nicht als Persönlichkeitsmerkmale gelten. Derartige Ansätze erschweren zum einen die Interpretation von Befunden, da unter der Annahme perspektivenspezifischer Validität von (Unterrichts-)Wahrnehmungen keine klaren Aussagen zum Zusammenhang von selbstwahrgenommener Lehrerpersönlichkeit (Lehrersicht) und fremdwahrgenommener Unterrichtsqualität (Schülersicht) abzuleiten sind (bzgl. a). Zum anderen lassen sich Befunde nicht vergleichen und verifizieren, sofern diese auf nicht-vergleichbaren Konstrukten und Erhebungsverfahren basieren, wie es vor allem in Bezug auf die Lehrerpersönlichkeit der Fall ist (bzgl. b); während mit den Prozessmerkmalen der Unterrichtsqualität aus Schülersicht mittlerweile ein weit etabliertes Konzept besteht, mit dem sich die Qualität von Unterricht gut abzeichnen und vergleichen lässt. Aktuellen Forschungsansätzen zur Bedeutung der Lehrerpersönlichkeit für die Unterrichtsqualität fehlt es demnach sowohl an Stringenz als auch an Vollständigkeit. An Stringenz fehlt es insofern, als dem aktuellen Forschungsstand zur perspektivenspezifischen Validität (vgl. Clausen 2002, Gruehn 2000) in Bezug auf die Lehrerpersönlichkeit kaum Rechnung getragen wird. Das Forschungsdesiderat der Vollständigkeit erfordert zudem ungekürzt durchdachte, differenzierte Konzepte wie es die Big-Five gegenüber separat und zusammenhanglos erhobenen Merkmalen darstellen; ebenso wie die Prozessmerkmale der Unterrichtsqualität gegenüber einzelnen globalen Einschätzungen. Um also sowohl stringent als auch möglichst vollständig die Bedeutung der Lehrerpersönlichkeit für die Unterrichtsqualität herauszustellen, lautet die zweite Forschungsfrage der vorliegenden Arbeit:

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten Lehrerpersönlichkeit und Unterrichtsqualität (aus Schülersicht)?

II. Methoden

6 Datengrundlage und Untersuchungsdesign

Die im Rahmen dieser Dissertation vorgenommenen Untersuchungen basieren auf Daten des DFG-Forschungsprojekts *Determinanten des Schülerfeedbacks*⁸². Im Folgenden werden daher Anliegen und Untersuchungsdesign der Hauptstudie knapp umschrieben. Daran anschließend werden die den Auswertungen dieser Arbeit zugrundeliegende Stichproben sowie das verwendete Erhebungsinventar samt seiner Validierung vorgestellt. Abschließend werden die Analyseverfahren zur Beantwortung der entwickelten Fragestellungen beschrieben.

Grundsätzlich sind bei empirischen Erfassungen theoretischer Gegenstände zwei Beobachtungsformen zu unterscheiden: Niedrig-inferente Beobachtungen sind spezifische Aspekte des beobachtbaren Verhaltens, die objektiv kodierbar sind und geringe Anforderungen an das Vorverständnis erfordern: Häufigkeit und/oder Dauer werden dabei direkt quantitativ erfasst. Hoch-inferente Messungen dagegen beziehen sich auf globale, abstrakte Aspekte und bedürfen der Interpretation. (Schüler-/Lehrer-)Befragungen mit Instrumenten der Klimaforschung gelten demnach als hoch-inferent, da Merkmale oberflächlich abgefragt werden (Clausen/Reusser/Klieme 2003). Erhebungen entlang von differenzierten Merkmalen – wie in der vorliegenden Untersuchung der Fall – stellen ein Mischverhältnis bzw. einen Kompromiss von Hoch- und Niedriginferenz dar.

6.1 Anlage der DFG-Studie Determinanten des Schülerfeedbacks

Das Forschungsanliegen dieser Studie liegt darin, Biasfaktoren von Schülerfeedbacks zur Unterrichtsqualität zu identifizieren. Schülerfeedbacks werden in der Regel fachübergreifend verglichen, was – so eine Hypothese der Hauptstudie – einen fairen Beurteilungsprozess der Unterrichtsqualität verhindert (vgl. hierzu auch Praetorius et al. 2014). Ziel der Studie ist es daher vor allem zu prüfen, ob Schülerfeedbacks als fachspezifisch oder fachübergreifend valide gelten können. Ein weiteres Ziel ist zu prüfen, ob sich differenzierte Einschätzungen des Unterrichts weniger verzerrt darstellen als globale Urteile. Darüber hinaus dient die

⁸² Gruehn, Rahn & Böttcher, DFG-Geschäftszeichen: „Determinanten des Schülerfeedbacks“ - RA 2380/1-1, GR 1951/2-1.

Studie dazu, die Forschungslücke zur schülerperzipierten Unterrichtsqualität von Schülerinnen und Schülern an gymnasialen Oberstufen von Gesamtschulen und Berufskollegs zu schließen. Dazu wurden Schülerinnen und Schüler der Jahrgangsstufe 11 von Gesamtschulen und Berufskollegs in ganz Nordrhein-Westfalen hinsichtlich verschiedener Unterrichtsfächer per Fragebogen befragt.

Geplant war, je zwanzig Berufskollegs mit technischem, wirtschaftlichem und erziehungswissenschaftlichem Schwerpunkt sowie zwanzig Gesamtschulen zu erheben, wobei jede Schule mit zwei Klassen vertreten sein sollte, was (bei einer Kalkulation von durchschnittlich 20 Schülerinnen und Schülern pro Klasse) einen Stichprobenumfang von ca. 3.200 Schülerinnen und Schülern sowie maximal 440 Lehrerinnen und Lehrern ergibt (siehe Tab. 3).

Tab. 3: Angestrebte Sollstichprobe der durchgeführten Studie.

		Schulen	Klassen	Schülerinnen und Schüler	Lehrkräfte
Berufskollegs	Technik	20	60	~2400	max. 360
	Wirtschaft	20			
	Erziehung	20			
Gesamtschulen		20	40	~800	max. 80
Insgesamt		80	160	~3200	max. 440

6.2 Datenerhebung

Im März 2013 wurden insgesamt 117 zufällig ausgewählte Gesamtschulen und Berufskollegs Nordrhein-Westfalens mit der Bitte um Teilnahme an der von der Deutschen Forschungsgemeinschaft (DFG) geförderten Studie *Determinanten des Schülerfeedbacks* angeschrieben. 54 Schulen mit schätzungsweise ca. 128 Klassen und ca. 2.500 Schülerinnen und Schülern sowie maximal 343 Lehrkräften erklärten sich bereit, an der Studie teilzunehmen. Die Berufskollegs unterteilten sich in folgende Bildungsgänge: neun Schulen mit dem Schwerpunkt *Erziehung und Soziales*, neun Schulen mit dem Schwerpunkt *Technik* und 22 Schulen mit den Schwerpunkten *Wirtschaft und Verwaltung*. Es wurden in allen Schu-

len Daten zu den Fächern Deutsch und Mathematik erhoben, in den Berufskollegs zusätzlich zum jeweils profilbildenden Fach, das je nach Bildungsgang variiert (z.B. Erziehungswissenschaft, Technik oder Betriebswirtschaftslehre).

Die Erhebung fand im Zeitraum von Juni bis Mitte Juli 2013 statt. Diese Terminierung wurde bewusst an das Schuljahresende gekoppelt, um einen möglichst vertieften Eindruck der zu erhebenden Facetten des Unterrichts sowie der Einschätzung der Lehrerpersönlichkeit zu ermitteln. Diesem Vorgehen liegt die Annahme zugrunde, dass die Schülerinnen und Schüler nach fast einem ganzen Schuljahr zum einen charakteristische Merkmale des Unterrichts und der Lehrkraft ausreichend gut kennen, um diesbezüglich reliable Aussagen treffen zu können. Zum anderen wird davon ausgegangen, dass sich im Laufe des Schuljahres innerhalb der Klassen ein Klassenklima bzw. kollektive Lernumweltwahrnehmungen gebildet haben, die Rückschlüsse auf klassenspezifische Tendenzen zulassen (vgl. vertiefend hierzu Gruehn 2000).

Da Befunde zu Reihenfolgeeffekten bislang uneinheitlich ausfallen (vgl. Unkelbach et al. 2014, Walter/Rose 2013, Hohensinn et al. 2011) wurde verzerrenden Reihenfolgeeffekten vorgebeugt. Die Fragebögen für die Schülerinnen und Schüler wurden daher in acht verschiedenen Versionen erstellt, die in der Reihenfolge der Fächerabfrage systematisch variierten: es gab zwei Versionen für die Gesamtschulen, einmal in der Reihenfolge Mathematik-Deutsch (\triangleq GS_MD) und einmal in der Reihenfolge Deutsch-Mathematik (\triangleq GS_DM) und sechs Reihenfolgevarianten für die drei Fächer der Berufskollegfragebögen (BK_DMP, BK_DPM, BK_MDP, BK_MPD, BK_PDM, BK_PMD)⁸³.

Für die Befragung der Lehrkräfte wurden jeweils ein Gesamtschul- und ein Berufskollegfragebogen, bezogen auf den Unterricht in der entsprechend befragten Klasse bzw. des entsprechend befragten Kurses, erstellt. Die Lehrerhebung erfolgte zeitlich zwar weitgehend separat von der Schülerhebung, um den Lehrkräften die größtmögliche Selbstbestimmung bezüglich des Zeitpunkts und der Dauer der Bearbeitung ihres Fragebogens zu gewähren; dennoch wurde per Codierung gewährleistet, dass die Lehrerinnen und Lehrer ihren entsprechenden Klassen zugeordnet werden konnten. Insgesamt 168 Lehrkräfte ließen uns ihre ausgefüllten Fragebögen zukommen.

⁸³ Im Text oder in Tabellen werden in Bezug auf die Fächer gelegentlich die Abkürzungen ‚D‘ für das Fach Deutsch und ‚M‘ für Mathematik verwendet.

6.3 Stichprobenbeschreibung der verwendeten Daten

In der vorliegenden Arbeit werden ausschließlich Daten der Fächer *Mathematik* und *Deutsch* ausgewertet ohne Differenzierung der Profilschwerpunkte der Berufskollegs. Diese sind in dieser Arbeit irrelevant. Im weiteren Verlauf wird zudem von einer Unterteilung in die beiden verschiedenen Schulformen abgesehen (Näheres hierzu in Kap. 7).

Außerdem ist zu beachten, dass je nach Berechnung unterschiedliche Stichproben – entweder nur die der Schülerinnen und Schüler oder die Synthese der Schülerinnen und Schüler mit den entsprechenden Deutsch- und/oder Mathematiklehrkräften – relevant sind und die Fallzahlen dementsprechend variieren. Es ergeben sich daher für die vorliegende Arbeit folgende relevante Stichproben:

6.3.1 Schülerstichprobe

Insgesamt wurden 1577 Berufskollegiaten sowie 680 Gesamtschülerinnen und -schüler befragt. In sieben Fällen konnten allerdings entweder keine Klassen zugeordnet werden oder die Fragebögen waren durchweg fehlerhaft (sie wiesen entweder Ankreuzmuster auf oder waren überhaupt nicht ausgefüllt). Desweiteren ließen sich die Fragebögen von einer Schule nicht eindeutig zuordnen, an zwei Schulen lagen die Antworten von zu wenigen (<10) Schülerinnen und Schülern vor und eine Schule nahm nachträglich ihre Teilnahme an der Befragung zurück. Die endgültige Nettostichprobe der Erhebung umfasst demnach 2250 Schülerinnen und Schüler aus 123 Klassen an 49 Schulen. In der folgenden Tabelle (Tab. 4) wird die Verteilung der Schülerschaft in ihrer Gesamtheit dargestellt.

Tab. 4: Netto-Schülerstichprobe der durchgeführten Studie.

	Schülerinnen und Schüler	Klassen	Schulen
Gesamtschulen	674	36	10
Berufskollegs	1576	87	39
Insgesamt (% m/w)	2250 (49/51)	123	49

Es ist hier zu bemerken, dass zwar viermal so viele Berufskollegs wie Gesamtschulen erhoben wurden, jedoch nur etwa doppelt so viele Klassen aus den Berufskollegs stammen, da die Gesamtschulen stets mit mehreren Klassen an der Erhebung teilgenommen haben. Hinsichtlich der gesamten Schülerzahl verringert sich dieses Verhältnis nochmals, da die Klassen der Gesamtschulen fast durchweg die stärkeren Klassengrößen aufweisen. Die durchschnittliche Klassengröße liegt bei 18,3 Schülerinnen und Schülern. Zudem ist die Verteilung von Schülerinnen und Schülern so gut wie ausgeglichen.

6.3.2 Lehrerstichprobe

Aufgrund von Zuordnungsschwierigkeiten zwischen Klasse und unterrichtender Deutsch- bzw. Mathematiklehrkraft⁸⁴ lassen sich für die vorliegende Arbeit lediglich 103 Lehrerfragebögen verwenden. Diese umfassen 53 Deutsch- und 50 Mathematiklehrerinnen und -lehrer. Die Tabelle 5 liefert eine Übersicht über die Verteilung der verwendeten Lehrerstichprobe:

Tab. 5: Netto-Lehrerstichprobe der durchgeführten Studie.

	Lehrerinnen und Lehrer	...im Fach Deutsch	...im Fach Mathe- matik
Gesamtschulen	23	10	13
Berufskollegs	80	43	37
Insgesamt	103	53	50
(% w/m)	(39/61)	(50/50)	(28/72)

Ersichtlich wird hier, dass zwar im Fach *Deutsch* das Geschlechterverhältnis ausgeglichen ist, in *Mathematik* jedoch der Anteil der Lehrer mit 72% den Anteil der Lehrerinnen mit 28% deutlich überwiegt.

⁸⁴ Die Zuordnungen zu den Klassen boten Probleme, da entweder die notwendige Codierung am Fragebogen entfernt wurde, die entsprechende Klasse nicht befragt wurde oder zwei Lehrkräfte Fragebögen im selben Fach einer Klasse ausfüllten und nicht ersichtlich wurde, welche Schülerinnen und Schüler von welcher Lehrkraft unterrichtet wurden.

6.3.3 Gemeinsame Stichprobe von Lehrern und Schülern

Nicht in allen Klassen und Fächern ließen sich Schülerinnen und Schüler ihren Fachlehrerinnen und -lehrern zuordnen. Da die Schülerdaten in einzelnen Fragestellungen (Forschungsfrage 1) dieser Arbeit mit den Lehrerangaben direkt verknüpft werden, soll hier die gemeinsame Stichprobe von Lehrenden und Lernenden vorgestellt werden. Da von 103 Lehrkräften, die sich insgesamt 80 Klassen zuordnen lassen, Fragebögen vorliegen, reduziert sich die Schülerstichprobe auf 2106 Schülerinnen und Schüler bei Fragestellungen, die eine gemeinsame Analyse von Lehrer- und Schülerdaten erfordern. Für den Großteil der Klassen liegen entweder Daten der Deutsch- oder der Mathematiklehrkraft vor; in 24 Klassen (20 BK- und vier Gesamtschulklassen) jedoch haben sowohl die Mathematik- als auch die Deutschlehrkraft ihre Selbsteinschätzungen eingereicht. Die höhere Anzahl an Lehrkräften als an vorhandenen Klassen erklärt sich also dadurch, dass pro Klasse zwei Lehrkräfte (einmal im Fach Deutsch und einmal im Fach Mathematik) erfasst wurden, die (Soll-)Stichprobe also insgesamt doppelt so viele Lehrerinnen und Lehrer als Klassen umfasst.

Es liegen demnach folgende Daten für die Synthese aus Lehrer- und Schülerdaten vor:

Tab. 6: Stichprobe der Synthese aus Netto-Schüler- und –Lehrerdaten.

	Schülerinnen und Schüler	Lehrerinnen und Lehrer	Klassen	Schulen
Gesamt- schulen	606	23	19	10
Berufskollegs	1500	80	61	37
Insgesamt (% D/M)	2106 (50/50)	103 (52/48)	80	47

Entsprechend der nahezu 50:50-Verteilung der erhobenen Lehrkräfte auf die Fächer *Deutsch* und *Mathematik* sind auch deren unterrichtete Schülerinnen und Schüler auf die beiden Fächer verteilt.

7 Umgang mit der Mehrebenenstruktur der Daten

Die Schachtelung sowie die hierarchische Struktur der Datengrundlage müssen in den Analysen berücksichtigt werden, da die vorliegende Datengrundlage designbedingt gegen Regeln der Unabhängigkeit verstößt: aufgrund der Ziehung von Schulen und Klassen sind die Schülerdaten innerhalb von Klassen nicht zufällig und nicht unabhängig voneinander; die individuellen Angaben der Schülerinnen und Schüler (Level 1) sind innerhalb der Kontextebene *Klasse* (Level 2) angesiedelt. Die Klassenebene wiederum befindet sich innerhalb der Ebene der Schule (Level 3) und die Schulebene innerhalb der Schulform (Level 4). Die Schulformen wurden im Studiendesign vorab festgelegt, während es sich bei den Schulen um Zufallsstichproben handelt. Da die Klassen innerhalb der Schulen wiederum nicht zufällig, sondern aufgrund ihrer Schulzugehörigkeit und die Schülerinnen und Schüler entsprechend aufgrund ihrer Klassenzugehörigkeit (und somit resp. ihrer Schulzugehörigkeit) gezogen wurden, sind diese Ebenen nicht als unabhängig anzusehen. Bezüglich der Schul- sowie der Schulformebene lässt sich allerdings kein Hinweis in der Literatur finden, der darauf hindeutet, dass von diesen Ebenen Effekte auf die Wahrnehmung der Lehrerpersönlichkeit zu erwarten sind.⁸⁵ Daher erscheinen die über die Klasse hinausreichenden Ebenen für die Fragestellungen dieser Arbeit irrelevant. Darüber hinaus sind der jeweilige Unterricht und die jeweilige Persönlichkeit, die hinter diesem Unterricht steht, explizit Mittelpunkt der Fragestellungen – schul- bzw. schulformspezifische Perspektiven würden das Ausmaß an Komplexität der Modelle, die zur Beantwortung der eigentlichen Fragestellungen erstellt werden, sprengen. Die Reduktion der Modellkomplexität zugunsten der Übersichtlichkeit der Ergebnisse stellt daher ein weiteres Argument gegen den Einbezug von mehr als zwei Ebenen dar.

Die Klassenebene allerdings wird im Rahmen der Berechnungen stets berücksichtigt, sofern von einem nennenswerten Unterschied zwischen dieser Ebene

⁸⁵ Hinsichtlich der Unterrichtsqualität zeigt sich in der Untersuchung von Schurig et al. (2016) zwar durchaus ein Schulzugehörigkeitseffekt in Form eines mittleren ICC 1 von 0,6; dieser liegt jedoch deutlich unter jenen mittleren ICC 1, die auf Effekte der Klassenzugehörigkeit hinweisen (=0,17).

und dem individuellen Antwortverhalten auszugehen ist. Darüber hinaus ist bezüglich der Angaben zur Unterrichtsqualität sowie zur Wahrnehmung der Lehrerpersönlichkeit vorauszusetzen, dass die Schülerinnen und Schüler entlang einer gemeinsam geteilten Realität im Sinne des Klassenklimas (siehe hierzu bspw. Gruehn 2000) antworten, so dass von Klasseneffekten auszugehen ist.

Generell wird in den Berechnungen dieser Arbeit die hierarchische Struktur der vorliegenden Daten (auf Level 1 und Level 2) berücksichtigt, sofern dies angesichts der Intra- und Interklassenkorrelationen (ICC 1 und ICC 2) angebracht bzw. zulässig erscheint. Der ICC 1 gibt dabei an, wie groß der Anteil der Varianz auf der Level 2-Ebene an der Gesamtvarianz ist. Bereits relativ geringe ICC 1-Werte rechtfertigen die Einbeziehung der zweiten Ebene in die Analysen: Eine Intraklassenkorrelation (ICC 1) von bspw. 0,1 stellt eine 10%-ige Varianzaufklärung des Antwortverhaltens allein durch die Klassenzugehörigkeit dar und sollte daher in Form einer Mehrebenenanalyse berücksichtigt werden (vgl. Eid/Gollwitzer/Schmitt 2013, Christ/Schlüter 2012). Der ICC 2 dagegen gibt die Varianz zwischen den Klassen resp. Kontexteinheiten an und dient als Reliabilitätsmaß der geteilten Wahrnehmungen. Hinsichtlich manifester Variablen gelten ICC 2-Werte ab 0,7 als akzeptable Ausprägung, zwischen 0,5-0,7 als marginale Ausprägung und unter 0,5 als schwach (vgl. Stapleton/Yang/Hancock 2016, Klein et al. 2000). Die jeweiligen Ebenen werden bei den Analysen entsprechend gekennzeichnet und begründet. U.a. mit dem Statistikprogramm Mplus ist es möglich, diese besondere Datenstruktur in die Berechnungen einzubinden und anhand von Mehrebenenanalysen individuelle (Level 1 resp. *within*) und kollektive bzw. kontextuelle (Level 2 resp. *between*) Effekte voneinander zu trennen und simultan sowie separat zu analysieren. Dabei wird die Gesamtvarianz der untersuchten Gegenstände resp. Variablen auf die verschiedenen Ebenen aufgeteilt (Christ/Schlüter 2012). Auf individueller Ebene (Level 1) hat dies eine Korrektur der Standardfehler entlang der (die Unabhängigkeit des individuellen Antwortverhaltens verletzenden Level 2-)Clusterstruktur, die sich im Fall dieser Arbeit durch die Klassenzugehörigkeit ergibt, zur Folge. Dabei wird von korrelierten Beobachtungen innerhalb der jeweiligen Kontexteinheiten ausgegangen, wodurch die effektive Stichprobengröße korrigiert wird und Standardfehler adäquat geschätzt werden können, was wiederum zur Vermeidung der Inflation von Signifikanzen und damit einhergehenden Fit-Werten innerhalb der Strukturgleichungsmodelle (siehe nächstes Kapitel) beiträgt (Hox 2010, Christ/Schlüter 2012, Eid/Gollwitzer/Schmitt 2013). Auf

Klassenebene (Level 2) können zudem Parameter anhand der Maximierung der Loglikelihood-Funktion geschätzt werden. Dies erfolgt mittels sogenannter „Sandwich-Schätzer“, einem Wechselspiel zwischen Individual- und Kontextebene (Muthén & Muthén 2012).⁸⁶ Grundlegende Gleichungen der und Vertiefungen in die Mehrebenenanalyse sind in der Literatur bereits umfassend beschrieben – detaillierte Informationen zur Mehrebenenanalyse finden sich u.a. bei Hox (1995, 2010), Geiser (2013), Muthén & Muthén (2012).

⁸⁶ Die erfolgt konkret über Random-Slopes und Random-Intercepts.

8 Umgang mit fehlenden Werten

Die Ursachen für fehlende Werte⁸⁷ sind vielfältig. Neben intentionalen, d. h. dem Studiendesign geschuldeten, fehlenden Werten gibt es in den meisten quantitativen Untersuchungen *echte*, das heißt nicht-intendierte fehlende Werte, die möglicherweise die vorhandene Datengrundlage verzerren und so zu fehlerhaften Analyseergebnissen führen können. Entsprechend ist zwischen folgenden fehlenden Werten zu unterscheiden:

- MCAR (*Missing Completely At Random*): Als solche werden Werte bezeichnet, die bspw. durch versehentlich fehlerhafte Dateneingabe oder zufällige Auslassung bei der Beantwortung des Fragebogens fehlen. Es bestehen demnach keine systematischen Zusammenhänge der fehlenden Werte mit anderen Variablen oder der Ausprägung der entsprechenden Variablen.
- MAR (*Missing At Random*): Ebenso wie MCAR-bedingte fehlende Werte sind MAR-bedingte fehlende Werte in weiten Teilen zufällig. Bei Kontrolle zusätzlicher beobachteter Variablen zeigt sich, dass nicht die Ausprägung der fehlenden Variable selbst, sondern ein Zusammenhang mit einer oder mehreren anderen erhobenen Variablen in Teilen ursächlich für die fehlende Angabe ist. Sie ergeben sich demnach aus einer Funktion des fehlenden Wertes mit mindestens einer Kovariate bzw. mindestens einem anderen beobachteten Wert.
- MNAR (*Missing Not At Random*): Das Fehlen solcher Werte ist nicht zufällig, sondern lässt sich vollständig entweder durch die Ausprägung der Variable selbst und/oder durch Zusammenhänge mit anderen beobachteten Variablen erklären und voraussagen bspw. durch gezieltes Nichtbeantworten von bestimmten Befragten Gruppen oder bestimmte Antwortmuster (sogenannte *patterns*).
- (vgl. Rubin 1976, Muthén & Muthén 2012)

MCAR-*Missings* wurden im Datensatz bereinigt, indem sie, sofern möglich, korrigiert oder andernfalls als fehlende Werte definiert wurden. MNAR-*Missings*

⁸⁷ Fehlende Werte werden in der Literatur häufig auch als *Missings* oder als *missing data* bezeichnet.

konnten in der vorliegenden Untersuchung nicht festgestellt werden.⁸⁸ Schwieriger gestaltet sich der Umgang mit MAR-*Missings*, bei denen Rückschlüsse bezüglich des Fehlens der Angaben im vorliegenden Datensatz weder ermittelt noch ausgeschlossen werden konnten.

Je nach Art des Ausfallmechanismus⁴ empfehlen sich verschiedene Arten des Umgangs mit fehlenden Werten. Eine grundsätzliche Möglichkeit besteht im (listen- oder paarweisen) Ausschluss der Fälle mit fehlenden Werten. Hier sei jedoch angemerkt, dass es sinnvoll ist, das Ausmaß an fehlenden Werten und den damit einhergehenden Datenverlust möglichst gering zu halten: zum einen, um die Datengrundlage nicht zu stark zu reduzieren, zum anderen, um die Datengrundlage nicht systematisch zu verfälschen. Ersteres hat den Nachteil, dass ein geringerer Stichprobenumfang weniger effiziente Parameterschätzungen innerhalb von Analysen bedingen kann, mit Letzterem können systematisch verzerrte Schätzungen einhergehen (Lüdtke/Robitzsch/Trautwein/Köller 2007). Eine andere Möglichkeit ist das Ersetzen fehlender Werte durch Mittelwerte. Dieses Verfahren schwächt allerdings die Genauigkeit der Varianz betroffener Merkmale, was zur verzerrten Schätzung (Unterschätzung) von Standardfehlern führen kann (Reinecke 2005).

Die Großzahl der Berechnungen dieser Arbeit gründet auf Strukturgleichungsmodellen und wurde mit Mplus (Version 7) durchgeführt. Angesichts der vorab aufgeführten Nachteile von anderen Verfahren zum Umgang mit fehlenden Werten wurden in diesen Berechnungen⁸⁹ die fehlenden Werte per FIML (*Full Information Maximum Likelihood*) ersetzt, sofern nicht alle Angaben zu einer Skala komplett fehlen. Bei dieser Prozedur werden die fehlenden Werte anhand der Schätzung, des Vergleichs und der daraus resultierenden Berechnung von Parametern ersetzt. Konkret wird dabei auf Grundlage der Kovarianzmatrix aller beobachteten Variablen eine *Maximum-Likelihood*-Schätzung der fehlenden Werte vorgenommen, wobei nach einer Kombination von Parametern gesucht wird, die

⁸⁸ Die Überprüfung der Antwortmuster erfolgte mit Mplus im Rahmen des Analysebefehls *,twolevel basic*.

⁸⁹ So die konfirmatorischen Faktorenanalysen, Korrelationen sowie Pfadanalysen/Strukturmodelle.

die Wahrscheinlichkeit der Kovarianzmatrix der beobachteten Werte maximiert.⁹⁰ Vorteilhaft bei diesem Verfahren sind effiziente Schätzer und Standardfehler, zudem bleibt der Stichprobenumfang erhalten. Ein Nachteil ist, dass schiefe Daten ebendiese Schätzer und Standardfehler möglicherweise (bedingt) verzerren können, da die FIML-Prozedur auf der Annahme einer Normalverteilung basiert (Reinecke 2005).

Bei jenen Berechnungen, die mit SPSS durchgeführt wurden⁹¹, wurde nur auf vollständige Datensätze zurückgegriffen. Da keine MNAR-Missings festgestellt wurden, kann die Gefahr einer systematischen Verzerrung durch die Eliminierung der fehlenden Werte weitgehend ausgeschlossen werden. Die Beschränkung hat – wie oben erwähnt – zwar dennoch den Nachteil, dass sie die Stichprobengröße reduziert. Allerdings dient dieses Vorgehen der Überprüfung der tatsächlich vorliegenden Datengrundlage, was hier vor allem bei der Validierung der für die Fragestellungen relevanten Skalen als wichtig erachtet wurde.

⁹⁰ Genauere Informationen zum Vorgehen finden sich bei Graham (2003).

⁹¹ Dies ist bspw. der Fall bei Reliabilitätsanalysen zur Ermittlung der Cronbachs α -Werte.

9 Interpretation der Modellgüte

Der Interpretation der Modellgüte liegen verschiedene Aspekte zugrunde. Neben der inhaltlichen Validierung von Konstrukten und entsprechenden Modellen sind statistische Kriterien sowie der *Modelfit* für die Einschätzung der Passung zwischen der hypothetischen (theoretischen) Modellstruktur und den tatsächlichen Beobachtungen (Datenbasis) bedeutsam. Zur Interpretation der Passung von Datengrundlage und statistischen Modellen bieten sich unterschiedliche Messwerte – sogenannte Fit-Indizes – an. Die jeweiligen Fit-Indizes basieren auf unterschiedlichen statistischen Messungen und verfügen demnach über unterschiedliche Vor- und Nachteile, aufgrund dessen es angemessen erscheint, mehrere Werte in die Interpretationen der Modellgüte einzubeziehen sowie die jeweilige Modellstruktur zu berücksichtigen. Das bedeutet auch, dass die Verletzung einzelner Fit-Indizes nicht sogleich zu einer Verwerfung des Modells führen sollte. Für die Einschätzung und Beurteilung der Modelle dieser Arbeit erscheint eine Kombination der im Folgenden mit ihren Stärken und Schwächen berichteten relevanten Fit-Indizes (vgl. Weiber/Mühlhaus 2010, Homburg/Baumgartner 1995) plausibel:

Im Chi-Quadrat-Test (auch χ^2 oder Chi^2) wird die Abweichung der beobachteten empirischen Kovarianzmatrix von der im Modell implizierten Kovarianzmatrix (gemäß der Nullhypothese) geprüft, wodurch gezeigt wird, inwiefern sich beobachtete Häufigkeiten von einer angenommenen Nullhypothese unterscheiden. Der χ^2 -Test schätzt demnach ebenso wie die meisten Fit-Indizes die Differenz zwischen einem Alternativmodell (in der Regel einem Nullmodell⁹²) und der vorhandenen Datenbasis. Dabei ist der χ^2 -Wert umso kleiner, je geringer die Differenz zwischen den Modellen ist. Ist das Ergebnis signifikant, entsprechen die empirischen Daten dem Nullmodell so gut, dass die Annahme von Effekten innerhalb des Modells verworfen werden kann (Zinnbauer/Eberl 2004). Ein großer Nachteil dieses Gütemaßes ist, dass seine Signifikanz von der Größe der Stichprobe und der Verteilungsform abhängt: während bei kleineren Stichproben die Unterschiede zwischen den Modellen kaum je Signifikanz ausweisen, ist dies bei großen Stichproben schon bei kleinsten Differenzen der Fall (Daniels 2008). Bei

⁹² In der Literatur wird dieses auch als Basismodell oder unabhängiges (*independence*) Modell bezeichnet.

großen Stichproben gehört demnach gemäß Chi²-Test nahezu jedes Modell verworfen (Zinnbauer/Eberl 2004). Durch den Einbezug der Freiheitsgrade (*df*=*Degrees of Freedom*) erhält die Stichprobenverteilung zwar jeweils eine andere Form, woraus sich schon eher eine *tatsächliche* Signifikanz (im Sinne klassischer Testtheorie) des Wertes ableiten und beurteilen lässt.

„Der χ^2 -Wert ist zwar nicht als inferenzstatistisches Gütekriterium brauchbar, jedoch nach Standardisierung mit der Zahl der Freiheitsgrade als deskriptives Anpassungsmaß sehr gut anwendbar (vgl. Jöreskog/Sörbom, 1982, S. 408). Von einem gut angepassten Modell kann dann ausgegangen werden, wenn das Verhältnis von χ^2 -Wert zu den Freiheitsgraden kleiner oder gleich 2,5 ist (vgl. Homburg/Baumgartner, 1995, S. 169).“ (Zinnbauer/Eberl 2004, S. 11)

Hierbei ist nun jedoch das Problem, dass die Aufnahme der Freiheitsgrade zur Berechnung der Modellgüte in Form des Bruchs Chi²/*df* eine hohe Parameterzahl bestraft, d.h. schon geringe Modellmodifikationen in Form zusätzlicher Parameter führen kaum zu einem besseren *Modelfit* (Daniels 2008). Darüber hinaus verfälschen bereits triviale Abweichungen von der Normalverteilung den χ^2 -Wert dahingehend, dass sich dieser unverhältnismäßig steigert (Daniels 2008).

„Um die Probleme des χ^2 -Maßes zu umgehen, empfehlen Browne/Cudeck (1993) die Verwendung des ursprünglich von Steiger/Lind (1980) entwickelten RMSEA (Root Mean Square Error of Approximation).“ (Zinnbauer/Eberl 2004, S. 10)

Der RMSEA basiert ebenfalls in hohem Maße auf der Approximation-bezogenen Diskrepanz⁹³ zwischen dem zu prüfenden und dem meist alternativen Nullmodell, ist jedoch eher unabhängig vom Umfang der Stichprobe sowie von der Verteilungsform. Da aber auch dieser Fit-Index die Freiheitsgrade impliziert, bestraft auch er Modellkomplexität (Zinnbauer/Eberl 2004, Daniels 2008) – ebenso wie der CFI (*Comparative Fit Index*), der sich als globales inkrementelles⁹⁴ Anpassungsmaß etabliert hat und als robustes Schätzmaß gilt (Weiber/Mühlhaus 2010, siehe hierzu auch Homburg/Baumgartner 1995). Allerdings impliziert der

⁹³ Diese an dem Näherungsverfahren orientierte Diskrepanz steht im Gegensatz zur auf die Schätzung bezogene oder die auf den Gesamtfehler bezogene Diskrepanz (Zinnbauer/Eberl 2004).

⁹⁴ *Inkrementell* bezeichnet in diesem Zusammenhang einen Prozess, in dem kleinschrittig die kontinuierliche Verbesserung des Modells geprüft wird.

CFI den χ^2 , wodurch er teilweise durch die Größe der Stichprobe beeinflusst wird (vgl. Bentler 1990). Daher wird hier ergänzend der TLI (*Tucker-Lewis-Index*) zur Interpretation des CFI hinzugezogen. Dieser stellt ebenfalls einen inkrementellen jedoch stichprobenunabhängigen Fit-Index dar. D.h. die Modellbeurteilung basiert auf dem Vergleich eines Idealmodells (saturiertes Modell) und eines Fehl-Modells (unabhängiges Nullmodell) unter zusätzlicher Beachtung der Freiheitsgrade (Weiber/Mühlhaus 2010). Das saturierte Modell weist mit einem Wert von 1 eine perfekte Modellpassung auf. Werte, die größer als 1 ausfallen, weisen auf eine Überidentifikation bzw. Fehlspezifikation des Modells hin. Das Nullmodell hingegen weist mit einem Wert von 0 die schlechteste Passung auf. Der SRMR (*Standardized Root Mean Square Residual*) hingegen stellt ein rein deskriptives Anpassungsmaß dar und bestraft die Modellkomplexität resp. die Parameteranzahl nicht, da er die Freiheitsgrade nicht impliziert. Seine Berechnung gründet auf der Residuendifferenz zwischen Teilen der empirischen und spezifizierten Kovarianzmatrix. Allerdings verbessert sich dieser Fit-Index mit der Größe der Stichprobe deutlich (Homburg/Baumgartner 1995) und wird daher lediglich in Kombination mit anderen Indizes interpretiert.

Um verschiedene Modelle direkt miteinander zu vergleichen, eignen sich relativ globale Fit-Indizes wie der AIC (*Akaike Information Criterion*):

„Mit Hilfe von Akaike’s Informationskriterium [...] wird aus einer Reihe von alternativen Modellen das Modell ausgewählt, welches bei Genauigkeitsgrad die geringste Komplexität aufweist [...]. Von mehreren alternativen Modellen ist dasjenige mit dem geringsten AIC zu bevorzugen (vgl. Homburg/Baumgartner 1995, S. 169).“ (Zinnbauer/Eberl 2004, S. 12)

Ebenso zieht der BIC (*Bayesian Information Criterion*)

„sparsame Modellspezifikationen gegenüber Modellen mit vielen, teils irrelevanten, Parametern vor (vgl. Raftery, 1993, S. 164 f.). Die Größe ist dabei eine Approximation für den sog. Bayes-Faktor. Werden zwei konkurrierende Modelle 0 M und 1 M bei gegebenem Datensatz D betrachtet, gibt er an, wie gut das Modell 0 M im Vergleich zum Modell 1 M die Daten vorhersagt. Nach Raftery (1995, S. 134) können mit BIC zwei Modelle leicht verglichen werden: Das Modell mit dem kleineren BIC-Wert wird bevorzugt. Der BIC- Wert kann neben der Berechnung in Relation zum saturierten Modell auch in Relation zu einem Nullmodell berechnet werden.“ (Zinnbauer/Eberl 2004, S. 12f)

Da keine eindeutigen bzw. universellen Maßstäbe für die Interpretation der Fit-Indizes vorliegen, sondern diese vor allem in Kombination sowie der Modellstruktur und Datengrundlage vorgenommen werden kann, liegen lediglich grobe Richtlinien gemäß *Daumenregel* (vgl. Hu & Bentler 1999) vor. Eine zusammenfassende Tabelle (Tab. 7) bietet eine Übersicht über die in dieser Arbeit zur Beurteilung statistischer Modelle herangezogenen Werte der jeweiligen Gütekriterien:

Tab. 7: *Fit-Indizes nach Schermelleh-Engel/Moosbrugger/Müller 2003, S. 38; Werte in Klammern nach Hu & Bentler 1999 (siehe auch Zinnbauer/Eberl 2004).*

	Guter Fit	Akzeptabler Fit	Besonderheiten
Chi²/df	0 ≤ Chi ² /df ≤ 2	2 < Chi ² /df ≤ 3	Abhängig von Parameteranzahl, Stichprobengröße und Normalverteilung
RMSEA	0 ≤ RMSEA ≤ ,05 (,06)	,05 (,06) ≤ RMSEA ≤ ,08	Abhängig von Parameteranzahl, stichprobenunabhängig
SRMR	0 ≤ SRMR ≤ ,05 (,08)	,05 (,08) ≤ SRMR ≤ ,1	Unabhängig von Parameteranzahl, stichprobenabhängig
CFI	,97 (,9) ≤ CFI ≤ 1	,95 (,80) ≤ CFI ≤ ,97 (,9)	Abhängig von Parameteranzahl, teilweise stichprobenabhängig
TLI	,97 ≤ TLI ≤ 1	,95 ≤ TLI ≤ ,97	Abhängig von Parameteranzahl, stichprobenunabhängig
AIC			Abhängig von Parameteranzahl
BIC	Im Vergleich zweier Modelle zeigen kleinere Werte die generell bessere Passung an		Abhängig vom Verhältnis der Parameterzahl und Stichprobengröße

Im Falle saturierter Modelle (wie bspw. bei manifesten Pfadmodellen ohne definierte Restriktionen sowie multiplen manifesten Regressionen) betragen sowohl χ^2 als auch die Freiheitsgrade Null, mit der Folge, dass die Indizes zur Beurteilung der Modellgüte nicht interpretierbar sind.

„Die Beurteilung der Modellgüte eines saturierten Modells konzentriert sich daher eher auf die geschätzten Modellparameter (insbesondere die geschätzten Pfadkoeffizienten) und die erzielte Varianzaufklärung (gemessen über die R²-Werte) als auf die von Mplus [...] ausgegebenen Modellfitindices.“
(Geiser 2011, S. 80)

Unabhängig von der Güte eines Modells werden demnach die einzelnen Modellparameter hinsichtlich ihrer Signifikanz entlang des *twotailed p-value* geprüft. Der *twotailed p-value* (im Weiteren als p-value bezeichnet) gibt an, mit welcher Wahrscheinlichkeit ein beobachteter Befund unter Annahme des Nullmodells auftritt. Ist der p-value kleiner als das zuvor festgesetzte Signifikanzniveau (α) kann das Nullmodell abgelehnt werden (siehe Blocker/Conway et al. 2006). Die Signifikanz vorgefundener Befunde wird demnach folgendermaßen eingeschätzt (vgl. Blocker/Conway et al. 2006, Erdfelder/Faul et al. 2010): In Abhängigkeit von der Stichprobengröße werden Ergebnisse auf Basis der Individualebene von Schülerinnen und Schülern ($N=2250$) ab einem $p\text{-value}<0,01$, entsprechend dem 1%-Niveau als signifikant erachtet. Auf Klassenebene und auch der Individual-ebene der Lehrkräfte hingegen sinkt die Stichprobengröße auf $N=123$ bzw. $N=103$, so dass für Ergebnisse auf diesen Ebenen ein 5%-Signifikanzniveau (also ein $p\text{-value}<0,05$) angesetzt wird (Geiser 2013). Eindeutig signifikante Befunde werden mit zwei Sternen (**) markiert, knapp signifikante Befunde⁹⁵ mit einem Stern (*) und solche, die eine Signifikanz knapp verfehlen⁹⁶ mit einem Plus (+). Die F-Werte⁹⁷ werden nur im Falle von signifikanten Mittelwertunterschieden⁹⁸ vollständigshalber berichtet. Zusätzlich zu den p-values werden die (standardisierten) Koeffizienten (β - resp. Beta-Gewichte) der (bivariaten) Korrelationen und (multiplen) Regressionen einbezogen, um die Stärke des Zusammenhangs bzw. die Effektgröße aufzuzeigen: ab einer Höhe von (gerundet) $r=0,3$ gelten diese als moderat, ab (gerundet) $r=0,5$ als stark (vgl. Cohen 1988, Bortz/Schuster 2010).

⁹⁵ Als *knapp signifikant* wird in dieser Arbeit ein Ergebnis interpretiert, wenn es zwischen einem halben (Individualebene) und einem ganzen (Klassenebene) Prozentpunkt unterhalb des angesetzten Signifikanzniveaus und dem angesetzten Signifikanzniveau liegt (also bei einem p-value zwischen ,005 und ,01 auf Individualebene und ,04 und ,05 auf Klassenebene).

⁹⁶ Diese werden angegeben, wenn das Ergebnis mit bis zu einem halben (Individualebene) bzw. einem (Klassenebene) Prozent das festgesetzte Signifikanzniveau knapp verfehlt.

⁹⁷ Der F-Wert ergibt sich aus der – unter Beachtung der jeweiligen Freiheitsgrade – Schätzung der Varianzen innerhalb und zwischen den untersuchten Gruppen.

⁹⁸ Diese wurden mit SPSS entlang des t-Tests (unabhängige Stichprobe, zweiseitige Signifikanz) durchgeführt.

10 Erhebungsinventar und Validierungsverfahren

Das für die vorliegende Arbeit relevante Erhebungsinventar bezieht sich auf die Konstrukte *Persönlichkeit* und *Unterrichtsqualität*. Die jeweiligen Instrumente implizieren – je nach Konstrukt und dessen Komplexität – unterschiedlich viele Skalen und Perspektiven: die Unterrichtsqualität wird entlang ihrer Prozessmerkmale mit sieben Skalen und die Lehrerpersönlichkeit entlang der BigFive-Dimensionen mit fünf Skalen erhoben. Zudem werden in Bezug auf das Persönlichkeitskonstrukt neben den Schülerantworten zur Fremdeinschätzung der Lehrerpersönlichkeit auch die Selbsteinschätzungen der Lehrerinnen und Lehrer betrachtet, während die Prozessmerkmale der Unterrichtsqualität ausschließlich aus Schülersicht erfragt wurden. Die entsprechenden wortgetreuen Passagen aus dem eingesetzten Fragebogen sind im Anhang beigefügt (Anhang A).

Das Instrument zur Operationalisierung der *Persönlichkeit* aus Schülersicht wird zunächst anhand einer explorativen Faktorenanalyse auf die Replikation der Big-Five-Struktur innerhalb der Schülerdaten überprüft. Dieser Schritt erscheint notwendig, da dieses Instrument bislang nicht hinsichtlich seiner Verwendung zur Fremdeinschätzung der Persönlichkeit validiert ist. Im Weiteren werden faktorielle Validitätsanalysen entlang einer Vier-Modell-Prüfung eingesetzt.

Da Erhebungen von Schülerwahrnehmungen generell und mit Bezug auf Unterrichtsqualität im Besonderen dem Verdacht eines verzerrten Antwortverhaltens unterliegen (vgl. Kap. 4.4.1), werden die einzelnen (ggf. gekürzten) Skalen der schülerperzipierten Konstrukte *Persönlichkeit* und *Unterrichtsqualität* anhand von verschiedenen Modellen auf ihre fachspezifische versus fachübergreifende Aussagekraft überprüft. Dies erfolgt per konfirmatorischer Faktorenanalysen (CFA) im Rahmen einer vierstufigen Modelltestung, der sogenannten Vier-Modell-Prüfung (siehe hierzu vertiefend Daniels 2008). Dazu werden verschiedene Strukturgleichungsmodelle der einzelnen Skalen erstellt. Diese Reihenfolge der Modelltestung erfolgt in dieser Arbeit entlang des Differenzierungsgrades, den das jeweilige Modell dem Antwortverhalten zugrunde legt.

- in einem ersten Modell wird demnach überprüft, ob das Antwortverhalten der Schülerinnen und Schüler fachübergreifend und ohne jegliche Unterscheidung zwischen den Fächern Mathematik und Deutsch erfolgt. Dazu

werden alle Antworten auf die Items einer Skala auf einem gemeinsamen Generalfaktor modelliert:

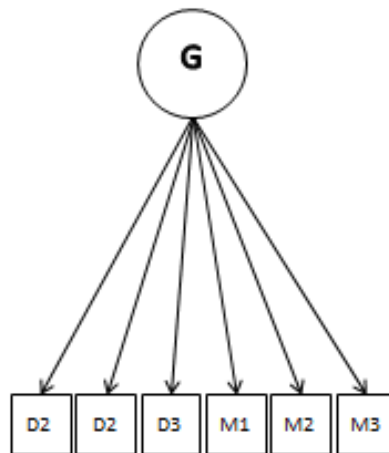


Abb. 4: Generalfaktormodell.

- Danach wird ein genestetes Modell erstellt, das die Antworten neben dem fachübergreifenden Generalfaktor gleichzeitig fachspezifisch in die Fächer Deutsch und Mathematik einordnet. Hier wird davon ausgegangen, dass die Schülerinnen und Schüler zwar durchaus zwischen den beiden Fächern differenzieren, jedoch gleichzeitig weiterhin einem beiden Fächern gemeinsamen Antwortverhalten folgen, das im Einsatz identischer Fragebögen in beiden Fächern, also durch einen methodischen Faktor begründet ist:

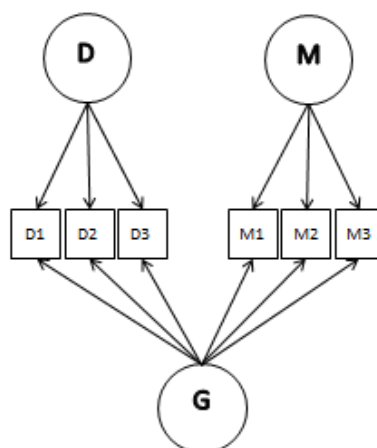


Abb. 5: genestetes Modell.

- Ein drittes Modell berücksichtigt ausschließlich die Fächertrennung und geht von einem fachspezifischen Antwortverhalten der Schülerinnen und Schüler aus. Dabei bestehen zwei Möglichkeiten:

- a) Die Fächer werden zwar getrennt modelliert, jedoch unter Zulassung von Korrelationen zwischen ihnen. Dieses Modell dient zur Überprüfung, ob sich trotz fachspezifischem Antwortverhalten auch Zusammenhänge zeigen, die Hinweise darauf geben, ob tatsächlich in beiden Fächern identische (hohe Korrelationen) oder unterschiedliche (schwache Korrelationen) Konstrukte gemessen wurden:

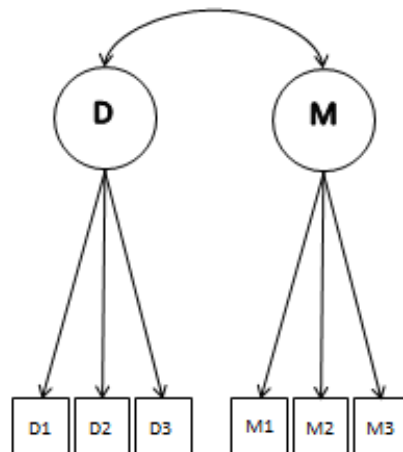


Abb. 6: fächergetrenntes Modell unter Zulassung von Korrelationen.

- b) Die Fächer werden vollständig getrennt modelliert. Dabei ist die Annahme, dass die Schülerinnen und Schüler vollständig zwischen den beiden Fächern Mathematik und Deutsch differenzieren:

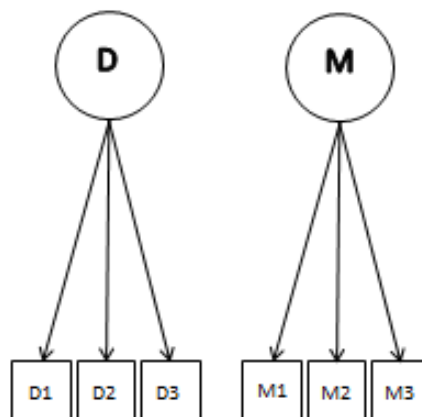


Abb. 7: fächergetrenntes Modell.

Dieser Analyseschritt wird mit einer Maximum-Likelihood-Methode abgeschlossen, bei der die Güte der verschiedenen Strukturgleichungsmodelle anhand ihrer Fit-Werte miteinander verglichen werden (siehe Kap. 9). Daraus lässt sich

die Struktur des in den Daten enthaltenen Antwortverhaltens ableiten, die für den Umgang mit den Daten in den weiteren Analysen – ob diese fachspezifisch oder fachübergreifend erfolgen – ausschlaggebend ist.

Im letzten Analyseschritt werden die Skalen dann zu latenten (ggf. modifizierten resp. gekürzten und/oder fachspezifischen) Messmodellen des jeweiligen Gesamtkonstrukts (*Persönlichkeit* und *Unterrichtsqualität*) zusammengeführt und per konfirmatorischer Faktorenanalysen mehrebenenanalytisch hinsichtlich ihrer Konstruktvalidität (gegenüber der vorangegangenen Skalenvalidität) betrachtet.⁹⁹ Innerhalb dieser Mehrebenenanalysen wird demnach zugunsten der Gesamtkonstrukte darauf verzichtet, die jeweiligen Dimensionen resp. Prozessmerkmale separat zu berichten, obgleich diese angesichts ihrer geringeren Komplexität bessere und deutlich robustere Model-Fit-Werte erwarten ließen. Allerdings würde ein Vorgehen der separierenden (und auch simplifizierenden) Datenanalyse entlang von Skalen anstatt von Konstrukten die angemessene Interpretation der Modelle in ihrer gesamten Komplexität zugunsten passender Werte vernachlässigen (vgl. Lang/Lüdtke 2005). Die Messmodelle werden daher entlang ihrer Konstruktstruktur als Gesamtmodell sowohl auf Individual- (Abb. 8) als auch auf Individual- und Klassenebene modelliert (Abb. 9) und einem Modell gegenübergestellt, das auf Klassenebene nur noch einen einzigen Generalfaktor berücksichtigt (Abb. 10).

⁹⁹ Da die Berechnung latenter Gesamtkonstrukte selbst für Mplus komplex ist, kann die gleichzeitige Modellierung von Skalen- und Konstruktvalidierung nicht berechnet werden. Die zuvor auf ihre Struktur geprüften separaten Skalen bilden demnach die Grundlage für die Modellierung der Skalen im Gesamtkonstrukt.

Erhebungsinventar und Validierungsverfahren

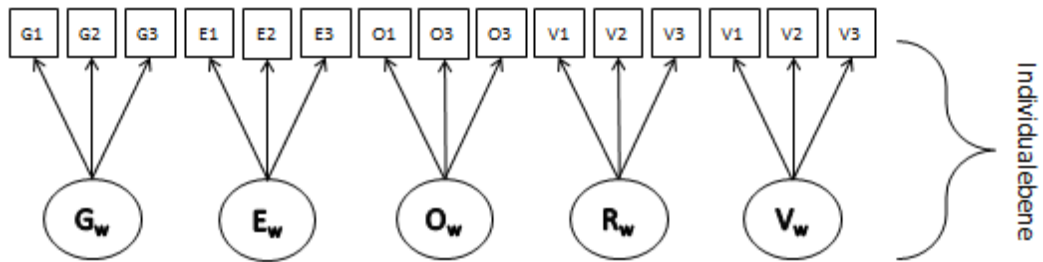


Abb. 8: Gesamtmodell (hier beispielhaft anhand der BigFive, ein adäquates Modell wird für die Prozessmerkmale des Unterrichts berechnet) auf Individualenebene.

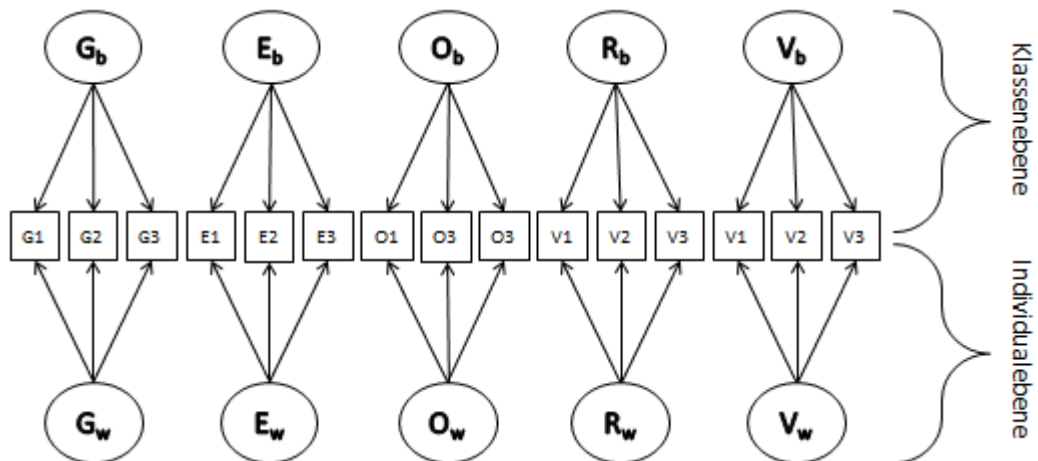


Abb. 9: Gesamtmodell (hier beispielhaft anhand der BigFive, ein adäquates Modell wird für die Prozessmerkmale des Unterrichts berechnet) auf Individual- sowie Klassenebene.

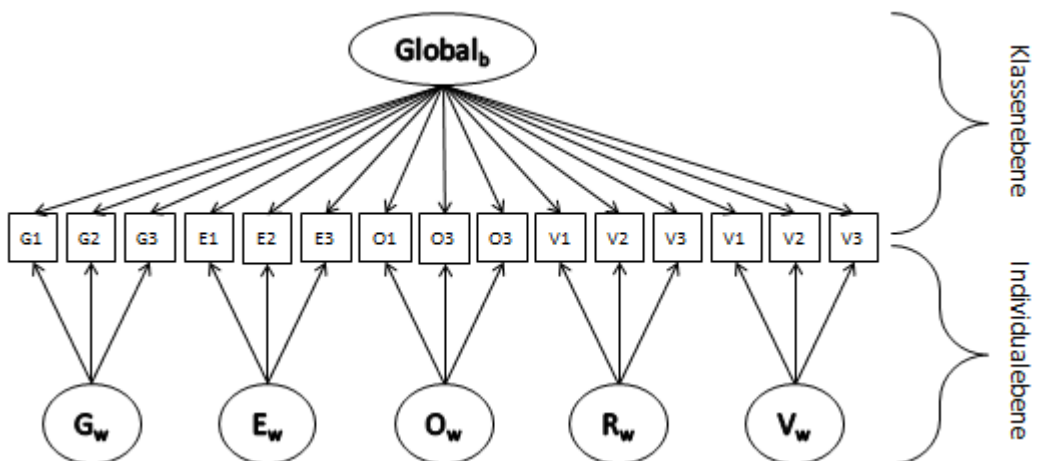


Abb. 10: Gesamtmodell (hier beispielhaft anhand der BigFive, ein adäquates Modell wird für die Prozessmerkmale des Unterrichts berechnet) auf Individualenebene und einem Generalfaktor auf Klassenebene.

Dieses kleinschrittige Verfahren dient der präzisen Berichterstattung der verwendeten Instrumente und soll es erleichtern, die Konstrukte in ihrer komplexen Operationalisierung nachzuvollziehen. Zudem gibt es einen vertieften Einblick in die Datenstruktur und klärt über implizite Varianzanteile auf. Dadurch wird ersichtlich, inwiefern die Antworten durch unterschiedliche Einflussfaktoren geprägt sind. Angesichts diverser Befunde, die belegen, dass Schülerinnen und Schüler durchaus deutlich zwischen Unterrichts- und persönlichen Lehrermerkmalen differenzieren (vgl. Kap. 2.2 und Feldmann 1986), wird dieser Aspekt in der vorliegenden Untersuchung nicht erneut aufgenommen.

Eher eine Auskunft bezweckt die Selbsteinschätzung der Lehrerpersönlichkeit. Hier wird allerdings explizit von fachspezifischen Analysen abgesehen, da es keine theoretischen Anhaltspunkte für die Annahme einer fachspezifischen Persönlichkeits-selbsteinschätzung gibt: Lehrkräfte unterrichten in der Regel mehrere Fächer und es ist kaum anzunehmen, dass sie relativ stabile eigene Persönlichkeitsmerkmale je nach Fach unterschiedlich an sich wahrnehmen. Angesichts dessen, dass das Instrument (BFI-S) hinsichtlich seiner Validität im Rahmen von Selbsteinschätzungen bereits gut geprüft ist, bedarf es keiner gesonderten Analysen.

Im Folgenden werden die Operationalisierungen der jeweiligen Konstrukte beschrieben und deren inhaltliche Validität überprüft. Die statistischen Reliabilitäts- und Validitätsprüfungen der Konstrukte innerhalb der erhobenen Daten werden jeweils im Anschluss daran berichtet.

11 Lehrerpersönlichkeit: Instrument

Die Persönlichkeit der Lehrkräfte wurde entlang des BigFive-Ansatzes operationalisiert: zum einen, weil dieses Konstrukt im Bereich der empirischen Persönlichkeitsforschung systematisch und gut einsetzbar ist; zum anderen, weil es in der wissenschaftlichen Diskussion weit verbreitet ist, was eine Vergleichbarkeit von Ergebnissen mit anderen Studien ermöglicht. Die *großen Fünf* dieses Konstrukts stehen für fünf grundlegende Eigenschaftsdimensionen, die sich aus jeweils sechs verschiedenen Facetten des menschlichen Verhaltensrepertoires zusammensetzen lassen und sich sowohl aus der klinisch-differentiellen Psychologie (Eysenck 1947) als auch ergänzend aus einem lexikalischen Vorgehen (Costa/McCrae 1987) herauskristallisiert haben (vgl. Kap. 3.3.1). Die fünf Dimensionen beziehen sich auf die *Gewissenhaftigkeit*, die *Extraversion*, die *Offenheit für (neue) Erfahrungen*, den *Neurotizismus* bzw. die *Resilienz* und die *Verträglichkeit* von Personen.¹⁰⁰ Anhand deren Systematik (siehe Kap. 3.3.2) wird ersichtlich, dass innerhalb der BigFive-Dimensionen der *Neurotizismus* die einzige Dimension ist, die nach ihrem – folgt man den Kriterien der sozialen Erwünschtheit – eher negativ attribuierten Pol benannt ist. Um in der vorliegenden Arbeit den eher negativen Charakter sowie das Alleinstellungsmerkmal dieser Dimension zu umgehen, wird im weiteren Verlauf dieser Arbeit die Dimension mit deren Gegenpol – der *Resilienz* – betitelt. Die Items der Skala wurden dementsprechend recodiert.

11.1 BFI-S zur Erhebung der Lehrerpersönlichkeit

Das in dieser Studie eingesetzte Instrument umfasst fünf durch das SOEP Berlin gut geprüfte, siebenstufige Kurzskalen zur Erhebung der jeweiligen BigFive-Dimension, den sogenannten BFI-S, der sich vor allem auf die *großen* Dimensionen konzentriert und die einzelnen Facetten eher sekundär behandelt (Gerlitz/Schupp 2005).¹⁰¹ Ein ausführlicheres Instrument innerhalb des deutschsprachigen

¹⁰⁰ Theoretische Grundlagen wurden im Theorieteil dieser Arbeit differenzierter bearbeitet (siehe Kap. 3).

¹⁰¹ Im Rahmen eines Pretests im Jahr 2004 wurde das dem BFI-S zugrundeliegende Inventar des BFI-25 (wie der Name schon sagt: mit 25 statt später im BFI-S mit 15 Items) anhand

Raums, das alle Facetten der entsprechenden Dimensionen in ihrer Breite berücksichtigt, ist der NEO-PI-R (von Borkenau/Ostendorf 1993), eine Kurzfassung des von Costa und McCrae Ende der 1980er entwickelten NEO-FFI, mit insgesamt 60 Items. Der Einsatz des dagegen mit 15 Items (es werden pro Dimension Facettenkombinationen mit jeweils einem Item erfragt¹⁰²) knapp gehaltenen BFI-S ist dem Umstand geschuldet, dass standardisierte schulische Fragebogenerhebungen immer innerhalb knapp bemessener Zeitrahmen stattfinden und zudem oft – wie auch in der dieser Teilstudie übergeordneten Erhebung – gleichzeitig mehrere Themenfelder bearbeiten. Obgleich tendenziell verkürzt, lässt sich die Struktur der BigFive mit diesem Instrument offenbar gut rekonstruieren (vgl. hierzu Schupp/Gerlitz 2005). Allerdings ist davon auszugehen, dass die interne Konsistenz der einzelnen Dimensionsskalen eher schwach ausgeprägt ist, da es sich a) um lediglich drei Items pro Dimension handelt, die zudem b) heterogene Facetten dieser jeweiligen Dimension zusammenfassen und nicht separat abbilden: „Interne Konsistenz wird somit zugunsten der Validität geopfert“ (Lang/Lüdtke 2005, S. 35). Zudem weisen einzelne Dimensionen möglicherweise erhöhte divergente Korrelationen auf, da sich die Konstrukte aufgrund ihrer hinsichtlich der Facetten tendenziell verkürzten Erfassung überlappen können, wie etwa die Dimensionen *Extraversion* und *Offenheit für (neue) Erfahrungen* (vgl. hierzu Lang/Lüdtke 2005). In der vorliegenden Studie sollte dennoch nicht gänzlich auf die Berücksichtigung der Facetten verzichtet werden, um trotz aller Kürze die BigFive inhaltlich möglichst breit abzubilden.¹⁰³

Die Konstruktgüte des BFI-S wurde lediglich – wie die meisten Instrumente zur Erhebung von Persönlichkeit (Lang/Lüdtke 2005) – in Bezug auf die Selbsteinschätzung von Personen überprüft, dies allerdings innerhalb einer breit gefächerten Altersklasse (repräsentative Zufallsstichprobe von Personen ab 16 Jahren, eine obere Altersbegrenzung ist in der einschlägigen Literatur nicht zu finden

verschiedener Kriterien systematisch gekürzt (vgl. Schupp/Gerlitz 2008). Die Facetten werden dabei nicht mehr separat, sondern zusammengefasst erfragt.

¹⁰² Um Antworttendenzen wie der Akquieszenz vorzubeugen, wurden – außer für die Dimension *Offenheit für (neue) Erfahrungen* – sowohl positiv als auch negativ formulierte Items eingesetzt, die für spätere Analysen entsprechend rekodiert wurden (vgl. Lang/Lüdtke 2005, Gerlitz/Schupp 2005).

¹⁰³ Zur Vertiefung bezüglich der *Kohärenzen* verschiedener und verschieden umfangreicher BigFive-Inventare bieten Lang und Lüdtke (2005) einen ausführlichen Überblick (vgl. auch Kap. 3).

(siehe Schupp/Gerlitz 2005)). Im Folgenden wird der in der SOEP-Erhebung eingesetzte BFI-S vorgestellt:

Ich bin jemand, der

- gründlich arbeitet
 - kommunikativ, geschwätzig ist
 - manchmal etwas grob zu anderen ist
 - originell ist, neue Ideen einbringt
 - sich oft Sorgen macht
 - verzeihen kann
 - eher faul ist
 - aus sich herausgehen kann, gesellig ist
 - künstlerische Erfahrungen schätzt
 - leicht nervös wird
 - Aufgaben wirksam und effizient erledigt
 - zurückhaltend ist
 - rücksichtsvoll und freundlich mit anderen umgeht
 - eine lebhaft Phantasie, Vorstellungen hat
 - entspannt ist, mit Stress gut umgehen kann
-

Skalierung von 1 (=trifft überhaupt nicht zu) über 2, 3, 4, 5, 6 bis 7 (=trifft voll zu)

Abb. 11: BFI-S gemäß SOEP (vgl. Gerlitz/Schupp 2005).

11.2 Beschreibung der einzelnen Skalen resp. Dimensionen

Die Skalen des als valide zu bezeichnenden Kurzinstrumentes zur Erhebung der Persönlichkeit im Rahmen des BigFive-Konstrukts (BFI-S) – geprüft und entwickelt innerhalb des Sozio-ökonomischen Panels (Dehne/Schupp 2007) – basieren zwar auf den fünf Persönlichkeitsdimensionen des BigFive-Konstrukts, erheben jedoch nicht explizit alle sechs heterogenen Facetten pro Dimension (=30 Facetten insgesamt), sondern sind entlang der Items mit den stärksten Ladungen

auf der jeweiligen Dimension entwickelt bzw. immer weiter reduziert worden.¹⁰⁴ Innerhalb des 2004 durchgeführten Pre-Tests des SOEP zur Entwicklung und Validierung des BFI-S-Inventars wurden dazu mittels Proxies¹⁰⁵ die Implikationen der heterogenen Facetten mit den verkürzt erhobenen BigFive-Dimensionen evaluiert. Proxies (Hilfsvariablen) bezeichnen in diesem Zusammenhang Items, die zur Erhebung der entsprechenden als eher latent zu bezeichnenden Facetten so formuliert sind, dass sie die Semantik der einzelnen Facetten sinngemäß und adäquat widerspiegeln. Dabei wurden vier der Facetten nicht explizit erhoben¹⁰⁶. Alle anderen zeigen signifikante Zusammenhänge zwischen den Proxies der Facetten und den zugehörigen Dimensionen bzw. Skalen des BFI-S, woraus sich schließen lässt, dass der BFI-S durchaus auch die komplexe Heterogenität der BigFive-Facetten weitgehend abbildet (Gerlitz/Schupp 2005).

¹⁰⁴ Ausgehend von einer Kurzbatterie des NEO-PI-R mit 44 Items wurde das Inventar erst auf 25 (BFI-25) und dann weiter auf die 15 Items des BFI-S gekürzt (Dehne/Schupp 2007).

¹⁰⁵ Eine Proxy-Variable misst solche Eigenschaften, die nicht direkt objektiv, reliabel, valide oder nicht mit vertretbarem Aufwand zugänglich sind. Im Fall der SOEP-Fragebogenentwicklung wurden die Proxies direkt entlang der Facettenbezeichnungen formuliert und die Skalen des BFI-S wurde in einem Pretest auf Korrelationen mit den Proxies der Facetten untersucht.

¹⁰⁶ *Herzlichkeit, Frohsinn, Offenheit für Gefühle, Bescheidenheit.*

Tab. 8: *In dieser Untersuchung eingesetztes Instrument (BFI-S): die Formulierungen aus Perspektive „Ich bin ein Mensch, der...“ bilden die originalgetreuen Formulierungen des BFI-S ab, während im Schülerfragebogen die Formulierungen an die Perspektive der Fremdeinschätzung angepasst wurden („Unser Lehrer/unsere Lehrerin...“).*

Dimension (Itemanzahl)	Beispielitem	
	Unser Lehrer/ unsere Lehrerin...	Ich bin ein Mensch, der...
Gewissenhaftigkeit (3)	erledigt Aufgaben wirksam und effizient.	Aufgaben wirksam und effizient erledigt.
Extraversion (3)	geht aus sich heraus, ist gesellig.	aus sich heraus geht, gesellig ist.
Offenheit für (neue) Erfahrungen (3)	hat eine lebhaft Phantasie, Vorstellung.	eine lebhaft Phantasie, Vorstellung hat.
Resilienz (resp. Neurotizismus) (3)	wird leicht nervös.	leicht nervös wird.
Verträglichkeit (3)	geht rücksichtsvoll und freundlich mit anderen um.	rücksichtsvoll und freundlich mit anderen umgeht.

Im Folgenden werden zunächst noch einmal die im BFI-S eingesetzten Formulierungen zur Erhebung der jeweiligen BigFive-Dimension einzeln betrachtet. In diesem Zusammenhang werden die Dimensionen anhand der Facetten, die ihnen jeweils zugeordnet sind, inhaltlich beschrieben bzw. deren konträren Pole attribuiert. Danach werden deren im SOEP-Pretest (2004) anhand der Proxies ermittelten Skalenkorrelationen dargestellt und abschließend die einzelnen im BFI-S eingesetzten Itemformulierungen innerhalb der Facetten semantisch eingeordnet. An dieser Stelle sei nochmals und explizit darauf hingewiesen, dass zwar jeweils einer der Ausprägungspole der BigFive repräsentativ für die jeweilige Dimension steht, jedoch beide Ausprägungsextreme dem Konstrukt implizit sind. Dasselbe gilt auch für die Subdimensionen bzw. Facetten der BigFive. Es handelt sich demnach lediglich um die Ausprägungen verschiedener Merkmale, die jedoch wertfrei zu betrachten sind. Die in den folgenden Beschreibungen kursiv

gesetzten Attributionen markieren demnach lediglich die zum Gegenpol der jeweiligen Dimension zugeordneten Formulierungen.

11.2.1 Gewissenhaftigkeit

Eingesetzte Attributionen:

eher faul

gründlich arbeitend

wirksame und effiziente Erledigung von Aufgaben

Die Persönlichkeitsdimension *Gewissenhaftigkeit* impliziert die Facetten *Kompetenz* (competence), *Ordnung* bzw. *Ordnungsliebe* (order), *Pflichtbewusstsein* (dutyfulness), *Leistungsstreben* (achievement striving), *Selbstdisziplin* (self-discipline) und *Besonnenheit*¹⁰⁷ (deliberation) (Angleitner/Ostendorf 2004). Diese Facetten werden in der Literatur anhand folgender Attribute operationalisiert:

- *Kompetenz* (starkes Selbstwertgefühl und die Überzeugung, fähig, umsichtig und effektiv zu handeln) – selbstüberzeugt, sicher, fähig und effektiv versus irritierbar, verunsicherbar
- *Ordnung* bzw. *Ordnungsliebe* (Ordentlichkeit, Systematik) – gut organisiert, systematisch, ordentlich versus unorganisiert, unmethodisch, chaotisch
- *Pflichtbewusstsein* (strikte Befolgung ethischer Prinzipien, Erfüllung moralischer Verpflichtungen) – gewissenhaft, zuverlässig versus flüchtig, unzuverlässig, locker
- *Leistungsstreben* (Fleiß, Zielstrebigkeit und Arbeitsfreude) – ehrgeizig, zielstrebig versus gleichgültig, schnell aufgebend
- *Selbstdisziplin* (Antriebsstärke, Fähigkeit Aufgaben trotz Langeweile oder Ablenkung zu Ende zu bringen) – Fokus auf Erledigung der Aufgabe, konzentriert versus nachlässig, zerstreut, ablenkbar
- *Besonnenheit* (vorsichtiges, umsichtiges Verhalten) – sorgfältig, unbeirrbar, nachhaltig versus hastig, planlos, spontan

(nach Simon 2006, S.7, in Klammern stehen die den Proxies zugrunde liegenden Attribute abgeleitet von Gerlitz/Schupp 2005 und Asendorpf/Neyer 2012).

¹⁰⁷ In anderen Beschreibungen wird diese Facette auch als Umsicht, Bedacht, Sorgfalt tituliert (bspw. Simon 2006).

Die eingesetzte Skala zur Erhebung der Dimension *Gewissenhaftigkeit* zeigte während ihrer Entwicklung im SOEP-Pretest die stärksten Korrelationen mit den Proxies der Facetten *Kompetenz*, *Pflichtbewusstsein* und *Selbstdisziplin* (Gerlitz/Schupp 2005). Während die Itemformulierungen „gründlich arbeitend“ und „wirksame und effiziente Erledigung von Aufgaben“ inhaltlich recht deutlich den Facetten *Pflichtbewusstsein* und *Kompetenz*, welche auch am höchsten mit der Skala korrelieren, zuzuordnen sind, lässt sich das Adjektiv „faul“ zwar durchaus der *Selbstdisziplin* im Sinne einer negativ gedrehten Fokussierung auf die Aufgabenbewältigung interpretieren, aber auch als Gegenpol zum *Leistungsstreben*.

11.2.2 Extraversion

Eingesetzte Attributionen:

zurückhaltend

kommunikativ, gesprächig

aus sich herausgehend, gesellig

Die Dimension *Extraversion* ist den Facetten *Herzlichkeit* (warmth), *Geselligkeit* (gregariousness), *Durchsetzungsfähigkeit* (assertiveness), *Aktivität* (activity), *Erlebnishunger* (excitement-seeking) und *Frohsinn* (positive emotions) übergeordnet (Angleitner/Ostendorf 2004).

Die Operationalisierung der Facetten erfolgt entlang folgender Beschreibungen:

- *Herzlichkeit* (Warmherzigkeit und Freundlichkeit) – herzlich, freundlich versus reserviert, formell
- *Geselligkeit* (Genuss der Gegenwart anderer) – gesellig versus zurückgezogen
- *Durchsetzungsfähigkeit* (dominantes und energisches Verhalten, Gefühl sozialer Überlegenheit) – bestimmt, durchsetzend versus zurückhaltend
- *Aktivität* (energie- und kraftvolles Auftreten) – hoher Grad an Aktivität versus weniger Aktivität, mehr Ruhe
- *Erlebnishunger* (Sehnsucht nach Stimulation) – Erregung und Aufregung liebend versus geringer Bedarf an Aufregungen, selbstgenügsam
- *Frohsinn* (optimistische Grundeinstellung) – heiter, fröhlich versus nüchtern, trocken

(nach Simon 2006, S.5, in Klammern stehen die den Proxies zugrunde liegenden Attribute abgeleitet von Gerlitz/Schupp 2005 und Asendorpf/Neyer 2012).

Hier korrelieren die Proxies aller Facetten nahezu gleichermaßen mit dieser Dimension des BFI-S, wobei die Korrelationen insgesamt schwächer als bei den anderen Dimensionen ausfallen (wenn auch trotzdem ebenfalls hochsignifikant) und die Facette *Frohsinn* nicht erhoben wurde (Gerlitz/Schupp 2005). Inhaltlich ist das Attribut „zurückhaltend“ innerhalb der Itemformulierung des BFI-S antagonistisch der Dimension *Durchsetzungsfähigkeit* zuzuordnen. „Kommunikativ, gesprächig“ sind hingegen Formulierungen, die sich in mehreren Facetten spiegeln können: der *Geselligkeit* ebenso wie der *Aktivität* als auch durchaus dem *Erlebnishunger* und der *Durchsetzungsfähigkeit*. Demgegenüber trifft die Beschreibung „aus sich herausgehend, gesellig“ recht eindeutig vor allem die Facette *Geselligkeit*, kann aber auch teilweise den *Erlebnishunger* betreffen. Auffällig ist, dass die Skala insgesamt eher auf soziale Aspekte der Extraversion ausgerichtet ist.

11.2.3 Offenheit für Erfahrungen

Eingesetzte Attributionen:

lebhaftes Phantasie, Vorstellung

künstlerisch, ästhetisch

originell, neue Ideen reinbringend

Die Facetten *Offenheit für Phantasie* (fantasy), *Ästhetik* (aesthetics), *Gefühle* (feelings), *Handlungen* (actions), *Ideen* (ideas) sowie eine *Offenheit des Werte- und Normensystems* (values) kennzeichnen die Dimension *Offenheit für Erfahrungen* (Angleitner/Ostendorf 2004):

- *Offenheit für Phantasie* (lebhaftes Vorstellungsvermögen und Kreativität) – imaginativ, kreativ, visionär versus pragmatisch, im Hier und Jetzt
- *Offenheit für Ästhetik* (Interesse für Kunst, Musik und Poesie) – Kunst (bspw. Musik, Malerei, Poesie) schätzend versus künstlerisch wenig interessiert, schnörkellos, ungekünstelt, einfach
- *Offenheit für Gefühle* (intensives Erleben von Glück und Unglück) – gefühlsintensiv, gefühlsbetont versus Gefühle ignorierend, trocken-sachlich
- *Offenheit für Handlungen* (Vielzahl an Hobbys, Ausprobieren verschiedener Aktivitäten) – Neues vorziehend, Vielfalt und Veränderung schätzend versus konservativ-konventionell, im Vertrauten verharrend
- *Offenheit für Ideen* (Verfolgen intellektueller Interessen als Selbstzweck) – intellektuell, abstrakt, spekulativ versus

- *Offenheit des Werte- und Normensystems* (kritisches Hinterfragen von sozialen, politischen und religiösen Werten) – offen für Relativität von Werten, liberal versus konservativ, traditionell

(nach Simon 2006, S.5, in Klammern stehen die den Proxies zugrunde liegenden Attribute abgeleitet von Gerlitz/Schupp 2005 und Asendorpf/Neyer 2012).

Offenheit für Phantasie, Ästhetik und Handlungen sind die Proxies der Facetten, mit denen die Dimension *Offenheit für Erfahrungen* am stärksten korreliert (Gerlitz/Schupp 2005). Während sich die Formulierung „lebhaftes Phantasie, Vorstellung“ in der Facette *Offenheit für Phantasie* und die Formulierung „künstlerische ästhetische Erfahrungen“ in der *Offenheit für Ästhetik* wiederfinden, zeigt sich in der Formulierung „originell, neue Ideen reinbringend“ sowohl eine *Offenheit für Ideen* als auch für *Handlungen*.

11.2.4 Resilienz resp. Neurotizismus

Eingesetzte Attributionen:

leicht nervös.

entspannt, guter Umgang mit Stress

oft Sorgen machend

Die *Resilienz* resp. der *Neurotizismus* als Persönlichkeitsdimension zeichnet sich durch die Facetten *Ängstlichkeit* (anxiety), *Reizbarkeit* (angry hostility), *Depression* (depression), *soziale Befangenheit* (self-consciousness), *Impulsivität* (impulsiveness) und *Verletzlichkeit* (vulnerability) aus (Angleitner/Ostendorf 2004).

Die bezüglich der Facetten zugeschriebenen Attributionen lauten:

- *Ängstlichkeit* (hohes Maß an Sorgen und Furcht) – ängstlich, besorgt versus unbesorgt, entspannt
- *Reizbarkeit* (Ärger, Frustration und Verbitterung) – erregbar, reizbar, frustriert versus ruhig, gelassen
- *Depression* (Niedergeschlagenheit) – pessimistisch versus optimistisch
- *soziale Befangenheit* (Scham, Verlegenheit und Unterlegenheitsgefühl) – befangen, gehemmt versus ungezwungen, unbefangen
- *Impulsivität* (Gefühl mangelnder Kontrolle von Begierden und Verlangen) – ungezügelt, exzessiv versus beherrscht, kontrolliert
- *Verletzlichkeit* (Stressanfälligkeit, Abhängigkeit und Panik) – verletzlich, vulnerabel versus stressresistent

(nach Simon 2006, S.4, in Klammern stehen die den Proxies zugrunde liegenden Attribute abgeleitet von Gerlitz/Schupp 2005 und Asendorpf/Neyer 2012).

Am stärksten korreliert mit dieser Dimension das Proxy der Facette *Depression*, alle weiteren zwar ebenfalls deutlich, jedoch deutlich schwächer (ähnlich der Dimension *Extraversion*). Am geringsten fällt in dieser Dimension die Korrelation mit dem Proxy der Facette *soziale Befangenheit* aus, die eher stärker (und negativ) mit der Dimension *Offenheit für Erfahrungen* zusammenhängt (Gerlitz/Schupp 2005). Über die Attribution *Depression*, die eher als übergeordnete Facette aufgefasst werden kann, hinaus lassen sich die Formulierungen präzisiert folgendermaßen einordnen: *Reizbarkeit* und *Verletzlichkeit* sind die Facetten, die sich semantisch in den Attributionen „entspannt, guten Umgang mit Stress“ spiegeln. „Oft Sorgen machend“ verweist hingegen auf *Ängstlichkeit*, während „leicht nervös“ neben der *Ängstlichkeit* primär in der Facette *Verletzlichkeit* Ausdruck findet.

11.2.5 Verträglichkeit

Eingesetzte Attributionen:

manchmal etwas grob zu anderen

rücksichtsvoll und freundlich mit anderen

verzeihend

In den Facetten *Vertrauen* (trust), *Freimütigkeit* (straightforwardness), *Altruismus* (altruism), *Entgegenkommen* (compliance), *Bescheidenheit* (modesty) und *Gutherzigkeit* (tender-mindedness) spiegelt sich die Dimension *Verträglichkeit* wider.

Diese Facetten werden in der Literatur anhand folgender Attribute untermalt:

- *Vertrauen* (Glauben an die Ehrlichkeit und die guten Absichten der Mitmenschen) – vertrauensvoll gegenüber anderen versus misstrauisch, vorsichtig gegenüber anderen
- *Freimütigkeit/Moral* (Offenherzigkeit, Aufrichtigkeit und Unbefangenheit) – aufrichtig, geradeheraus versus ‚bedeckt‘ haltend, vorsichtig
- *Altruismus* (Hilfsbereitschaft, Besorgnis um das Wohlergehen anderer) – selbstlos versus egozentrisch
- *Entgegenkommen* (Sanftheit, Nachgiebigkeit, die Neigung Aggressionen zu unterdrücken) – kooperativ, entgegenkommend, anpassend versus antagonistisch, aggressiv, kompetitiv

- *Bescheidenheit* (bescheidene Zurückhaltung) – bescheiden versus überlegen, aufdringlich
- *Gutherzigkeit* (Sympathie für und das Bemühen um das Wohlergehen anderer) – mitfühlend, teilnehmend versus distanziert, zurückhaltend

(Simon 2006, S.6, in Klammern stehen die den Proxies zugrunde liegenden Attribute abgeleitet von Gerlitz/Schupp 2005 und Asendorpf/Neyer 2012).

Die Proxies der Facetten *Gutherzigkeit* und *Freimütigkeit* korrelieren mit der Dimension *Verträglichkeit* am stärksten. Zudem lädt das Proxy der Facette *Vertrauen* stärker auf der Dimension *Resilienz* und auch das Proxy des *Altruismus* weist hohe Nebenladungen in allen anderen Dimensionen bis auf die *Resilienz* auf (Gerlitz/Schupp 2005). Inhaltlich steht die Formulierung „manchmal etwas grob zu anderen“ konträr zu den Facetten *Gutherzigkeit* und *Entgegenkommen*, kann jedoch auch *Freimütigkeit* widerspiegeln. „Rücksichtsvoll und freundlich mit anderen“ zu sein kann semantisch ebenfalls mehreren Facetten zugeordnet werden; so etwa dem *Vertrauen*, dem *Altruismus* und dem *Entgegenkommen* ebenso wie der *Gutherzigkeit*. Gleichmaßen verhält es sich bezüglich des Passus „verzeihend“.

11.3 Validierung des Persönlichkeitskonstrukts: Messmodelle der Schülerantworten

Die Skalen zur Erfassung der Lehrerpersönlichkeit aus Schüler- und aus Lehrerperspektive werden im Folgenden getrennt voneinander validiert. Während die Passung der Schülerantworten hinsichtlich des eingesetzten Instruments (BFI-S) gründlich, d.h. hinsichtlich ihrer Fachspezifität und Skala für Skala, überprüft wird, da das Instrument in Bezug auf die Fremdeinschätzung nicht validiert ist und möglicherweise fachspezifischen Biasfaktoren unterliegt, kann die Struktur des BFI-S innerhalb der Lehrerantworten fachübergreifend übernommen und validiert werden, da nicht von einer fachspezifischen Selbsteinschätzung der Lehrkräfte auszugehen ist.

Zunächst wird anhand einer explorativen Faktorenanalyse überprüft, ob sich die Daten der Schülerinnen und Schüler entlang der BigFive-Struktur abbilden lassen. Dieser Schritt erscheint erforderlich, da es sich bei diesen Daten um die Fremdeinschätzung der Lehrerpersönlichkeit handelt und das eingesetzte Instrument (BFI-S) bislang nicht in Bezug auf dieses Vorgehen validiert wurde.

Daran anschließend werden die fünf Skalen, die die fünf Dimensionen des Big-Five-Konstrukts im Erhebungsinventar abbilden, anhand vier verschiedener Messmodelle entlang einer Vier-Modell-Prüfung hinsichtlich einer möglichen Fachspezifität im Antwortverhalten der Schülerinnen und Schüler mittels konfirmatorischer Faktorenanalysen miteinander verglichen (siehe hierzu Kap. 11.3 und 12.3).

Nach dieser Überprüfung und gegebenenfalls Modifikation einzelner Dimensionen werden vollständigshalber die ermittelten Cronbachs α , ICC 1 und ICC 2 berichtet. Abschließend werden die fachspezifischen Messmodelle konfirmatorisch jeweils für Deutsch und Mathematik parallel in ihren Gesamtkonstrukten sowie mehrebenenanalytisch unter Einbezug der Klassenebene auf ihre Modellgüte überprüft.

11.3.1 Explorative Faktorenanalyse (EFA)

Da das Instrument des BFI-S bislang nicht für den Einsatz in der Fremdeinschätzung geprüft ist¹⁰⁸, wird im Rahmen der Validierung der Schülereinschätzungen zur Persönlichkeit ihrer Lehrkräfte in einem ersten Schritt getestet, ob das Instrument die Struktur der BigFive widerspiegelt. Dazu wird fachübergreifend eine explorative Faktorenanalyse (EFA) unter Vorgabe einer schrittweisen Erhöhung von eins bis fünf oblique rotierenden¹⁰⁹ Faktoren mit Mplus 7 durchgeführt. Gegenüber den ein- bis vierfaktoriellen Modelllösungen weist das fünffaktorielle Modell die besten Model-Fit-Werte auf und entspricht demnach am besten den erhobenen Daten. Dass die BigFive-Struktur grundsätzlich erhalten bleibt, lässt sich an der explorativen Faktorenanalyse in folgender Tabelle veranschaulichen:

¹⁰⁸ Eine Fremdeinschätzung gilt lediglich generell bei den BigFive als möglich.

¹⁰⁹ Hier wurden die Faktoren in Mplus 7 anhand einer gemischten (geomin) Rotationsmethode geschätzt, da sich die Angaben allesamt auf den Aspekt *Persönlichkeit* beziehen und demzufolge von einer Interkorrelation der Angaben ausgegangen werden muss, wie vorab bei der Beschreibung des Messinventars ebenfalls erläutert wurde.

Lehrerpersönlichkeit: Instrument

Tab. 9: Explorative Faktorenanalyse der Lehrer-BigFive aus Schülersicht: Faktorladungen $\leq 0,2$ werden nicht berichtet; $N=4481$ (listenweiser Fallausschluss: 19 Fälle lassen Missings auf allen erhobenen Variablen verzeichnen). Die jeweils negativ gepolten Items sind kursiv gesetzt.

Unsere Lehrerin/ unser Lehrer...	Faktor 1 Gewissenhaftigkeit	Faktor 2 Extra-version	Faktor 3 Offenheit	Faktor 4 Resilienz	Faktor 5 Verträglichkeit
arbeitet gründlich	,91				
erledigt Aufgaben wirksam und effizient	,86				
<i>ist eher faul</i>	,44				
kommunikativ, gesprächig		,68			
geht aus sich heraus, ist gesellig		,82			
<i>ist zurückhaltend</i>		,48		,22	-,20
hat eine lebhaft Phantasie, Vorstellung		(,37)	,54		
schätzt künstlerische ästhetische Erfahrungen			,82		
ist originell, bringt neue Ideen rein	,25		,54		
ist entspannt, kann mit Stress gut umgehen				,33	,28
<i>wird leicht nervös</i>				,85	
<i>macht sich oft Sorgen</i>	-	-	,22	,25	-

Lehrerpersönlichkeit: Instrument

Unsere Lehrerin/ unser Lehrer...	Faktor 1 Gewissen- haftigkeit	Faktor 2 Extra- version	Faktor 3 Offenheit	Faktor 4 Resilienz	Faktor 5 Verträglichkeit
geht rücksichtsvoll und freundlich mit anderen um					,85
kann verzeihen					,59
<i>ist manchmal etwas grob zu anderen</i>					,69
Chi ² /df = 323,20/40	CFI/TLI = ,99/,97	RMSEA = ,04	SRMR = ,01		

Lediglich die *Resilienz*-Dimension bildet eher mit zwei statt drei Items einen eigenen Faktor, wenn auch angesichts einer schwachen Faktorladung des Items mit der Formulierung „*Unser/e Lehrer/in ist ein Mensch, der entspannt ist, mit Stress gut umgehen kann*“. Das dritte Item dieser Skala („*Unser/e Lehrer/in ist ein Mensch, der sich oft Sorgen macht*“) wird im Weiteren ausgeschlossen, da seine Ladungen mit 0,25 auf Faktor 4 (der Dimension *Resilienz*) und einer Nebenladung von 0,22 auf Faktor 3 (der Dimension *Offenheit für neue Erfahrungen*) insgesamt schwach ausgeprägt (<0,3) sind. Da diese Skala insgesamt offenbar nicht ganz problemlos funktioniert, ist anzunehmen, dass möglicherweise gewisse neurotizistische Züge von Lehrkräften – wie etwa *Ängstlichkeit*, die sich vor allem in der Itemformulierung, ob sich eine Lehrkraft *oft Sorgen macht* (*Resilienz* 3) niederschlägt sowie etwas abgeschwächter der Umgang der Lehrkraft mit *Stress* (*Resilienz* 1) – für deren Schülerinnen und Schüler nicht eindeutig beobachtbar und zuzuordnen sind. Lediglich die Nervosität der Lehrkraft wäre demnach eindeutig von deren Schülerinnen und Schülern identifizierbar und von anderen faktoriellen Einflüssen deutlich abgrenzbar. Hinsichtlich der Befunde von Hefstee (1994) und Vazire (2010) bilden Fremdbefunde vor allem typische Verhaltensmuster ab, die in sozialen Interaktionen sichtbar auftreten. Insofern ist plausibel, dass die Dimension *Resilienz* generell als eher schlechter zu beobachtendes und daher schlechter von außen einzuschätzendes Merkmal gilt (vgl. Kap. 3.4.1). *Ängstlichkeit* vor allem in Form von *Sorgen*, teilweise auch in Form von *Stress* sind so gesehen eher Aspekte, die sich introspektiv erfassen lassen, während das Item mit der eindeutigsten Faktorenladung – „*Unser/e Lehrer/in ist*

ein Mensch, der leicht nervös wird“ – auf ein durchaus gut von außen zu beobachtendes und zuschreibbares Merkmal verweist.

Weiterhin ist zu beobachten, dass ein Item („*Unsere Lehrerin/unsere Lehrerin geht aus sich heraus, ist gesellig*“) der Dimension *Offenheit für Erfahrungen* außerdem hoch auf der Dimension *Extraversion* lädt. Derartige Abweichungen sind jedoch angesichts der weiter oben beschriebenen Heterogenität des Gesamtkonstrukts und unter dem Einbezug möglichst vieler Facetten der jeweiligen Dimension im Erhebungsinventar zu erwarten gewesen. Die Model-Fit-Werte zeigen allerdings an, dass die Faktorenstruktur insgesamt gut die empirischen Daten widerspiegelt.

11.3.2 Vier-Modell-Prüfung

Im Weiteren wurde nun die Fachspezifität bzw. -unabhängigkeit der von den Schülerinnen und Schülern eingeschätzten Lehrermerkmale per konfirmatorischer Faktorenanalysen entlang dreier Messmodelle auf Individualebene überprüft: In einem ersten Modell wird ein singulärer, fachübergreifender Generalfaktor angenommen (vgl. Abb. 4 in Kap. 10), dann zweitens ein genestetes Modell¹¹⁰ analysiert. Im Messmodell 3a wird die Fächerstruktur getrennt nach Mathematik und Deutsch für die jeweilige Skala modelliert, wobei Interkorrelationen zwischen den fachspezifisch ausgerichteten Merkmalen zugelassen werden (vgl. Abb. 6); und Messmodell 3b wiederum trennt die Fächer konsequent ohne Zulassung von Korrelationen zwischen den Fächern (vgl. Abb. 7). Dieses Vorgehen soll es ermöglichen, den Einfluss des Fachs auf das Antwortverhalten der Schülerinnen und Schüler sowie deren individuelle Urteilstendenzen abzuwägen.

Da sich die Dimension *Resilienz* mit lediglich zwei Items nicht identifizierbar ist, wird die Skala in diesen Analyseschritt nicht einbezogen. Im weiteren Verlauf dieser Arbeit wird demnach die Dimension *Resilienz* lediglich in den Strukturmodellen berücksichtigt, die auf manifesten Skalenmittelwerten beruhen, nicht jedoch in den differenzierteren latenten Messmodellen.

¹¹⁰ Ein Modell gilt dann als *genestet*, wenn es sowohl die theoretisch abgeleitete Faktorenstruktur als auch einen Generalfaktor annimmt (s. Kap. 10, Abb. 5).

Lehrerpersönlichkeit: Instrument

Tab. 10: Messmodelle der Persönlichkeitsdimensionen (N=2239-2242).

BigFive-Dimension	Messmodell	Chi2/df	CFI	TLI	RMSEA	SRMR	AIC/BIC
Gewissenhaftigkeit	1	4246,91/9	,21	-,32	,46	,21	46912,86/ 47015,78
	2	19,68/3	1	,98	,05	,01	44354,45/ 44491,67
	3a	20,32/8	1	1	,03	,02	44351,38/ 44460,01
	3b	44,90/9	,99	,99	,04	,02	44390,36/ 44493,28
Extraversion	1	1637,11/9	,38	-,03	,28	,18	48107,17/ 48210,09
	2	9,63/3	1	,99	,03	,02	46566,06/ 46703,29
	3a	34,48/8	,99	,98	,04	,02	46590,37/ 46699,00
	3b	36,71/9	,99	,98	,04	,04	46602,48/ 46705,40
Offenheit	1	2469,06/9	0	-,81	,35	,19	50610,97/ 50713,89
	2	20,48/3	,99	,96	,05	,03	48948,97/ 49086,20
	3a	29,00/8	,99	,98	,03	,02	48954,95/ 49063,59
	3b	45,05/9	,98	,97	,04	,06	48992,66/ 49095,58
Verträglichkeit	1	4494,61/9	0	-,71	,47	,19	50597,43/ 50700,34
	2	59,91/3	,99	,94	,09	,06	48835,18/ 48972,40
	3a	59,88/8	,99	,98	,05	,03	48845,35/ 48953,98
	3b	78,51/9	,98	,97	,06	,07	48888,44/ 48991,35

Anmerkung: Die Modelle mit den besten Model-Fit-Indizes sind grau unterlegt.

Es wird ersichtlich, dass das fachspezifische Modell 3a, das Korrelationen zwischen den Fächern zulässt, fast durchweg (außer bei der *Extraversion*) leicht bessere Model-Fit-Werte aufweist als das genestete Modell (2). Dies weist auf einen methodischen Biaseffekt innerhalb differenzierter (und auch fachspezifischer) Einschätzungen der Schülerinnen und Schüler hin, der sich durch den Einsatz identischer Instrumente erklären lässt. Da die Fit-Indizes zwischen den Messmodellen 2, 3a und 3b marginal verschieden sind, während das Generalfaktormodell (1) die Datenstruktur sehr schlecht spiegelt, lässt sich insgesamt ableiten, dass das Fach stärker wirkt als die individuellen Urteilstendenzen. Zur präziseren Einschätzung werden zudem die Korrelationen zwischen den Fächern herangezogen:

Tab. 11: *Deutsch-Mathematik-Korrelationen (manifest auf Individualebene) innerhalb der einander entsprechenden Dimensionen (Signifikanz auf 0,01-Niveau).*

Gewissenhaftigkeit	Extraversion	Offenheit	Resilienz	Verträglichkeit
,14**	,07	,13**	,03	,15**

Da die Fachinterkorrelationen vernachlässigend gering ausfallen, ist hier gemäß den Regeln empirischer Forschung (vgl. Raftery 1993, Daniels 2008) das einfachste – d. h. das rein fachspezifische – Modell (3b) für weitere Berechnungen zu wählen.

11.3.3 Reliabilität der einzelnen Skalen resp. Dimensionen

Die demzufolge erstellten fachspezifischen Skalenwerte der klassischen Testgüte (Cronbachs α) sowie die Intra- und Interklassenkorrelationen (ICC 1 und ICC 2) und Model-Fit-Indizes (χ^2 , CFI/TLI, RMSEA, SRMR) des für die einzelnen Dimensionen verwendeten – und im Falle der *Resilienz*-Dimension modifizierten – Inventars stellen sich abschließend für die beiden Fächer wie folgt dar:

Tab. 12: Skalenwerte der erfassten BigFive-Dimensionen (die Resilienzskala wurde um ein Item gekürzt; $N=2183-2227$).

BigFive-Dimension (Itemanzahl)	Cronbachs α		ICC 1		ICC 2	
	Deutsch	Mathe	Deutsch	Mathe	Deutsch	Mathe
Gewissenhaftigkeit (3)	,81	,79	,26	,32	,87	,90
Extraversion (3)	,71	,74	,27	,41	,87	,93
Offenheit (3)	,79	,77	,25	,29	,86	,89
Resilienz (2)	,58	,66	,22	,33	,83	,88
Verträglichkeit (3)	,75	,78	,26	,40	,87	,93

Insgesamt zeichnen sich die Reliabilitätsmaße anhand des Cronbachs α eher schwach ab gegenüber den von Mayr und Neuweg (2006) gefundenen Cronbachs α , die über die per NEO-PI-R erhobenen Dimensionen durchschnittlich zwischen 0,87 und 0,97 variieren (vgl. in Kap. 2.2.3 erster Absatz). Die dort gemessene höhere Reliabilität bei den Schülerinnen und Schülern lässt sich zum einen durch die weitaus höhere Anzahl erfasster Items (120 insgesamt) erklären: mit vier Items pro Facette, umfasst die Skala einer Dimension 24 Items – gegenüber drei Items pro Dimension in der vorliegenden Arbeit. Die Auswirkung einer solch deutlichen Abweichung in der Skalenlänge auf den Cronbachs α lässt sich bspw. bei Gruehn (2000, S. 128) nachvollziehen. Zum anderen lassen sich die abweichenden Werte womöglich durch die Aggregation der Schülerantworten auf Klassenebene erklären, während die Reliabilität in dieser Arbeit bislang lediglich auf Individualebene erfolgt, wobei ICC 1 und ICC 2 deutlich auf eine geteilte Wahrnehmung hinweisen. Die Aggregation mehrerer Fremdeinschätzungen¹¹¹ ermöglicht die Reduktion von Urteilsverzerrungen sowie die Steigerung der Reliabilität von erhobenen Merkmalen im Sinne der internen Konsistenz (Cronbachs α , Intraklassenkorrelationen) (Wirtz 2014).

¹¹¹ Eine Aggregation erfordert mindestens vier individuelle Urteile (nach Wirtz 2014).

Erst infolge der insgesamt hohen Intraklassenkorrelationen (ICC 1), die auf eine Varianzaufklärung des Antwortverhaltens durch die Klassenzugehörigkeit zwischen 21,5% bis 40,6% hinweisen und im Fach *Mathematik* durchweg ausgeprägter sind als im Fach *Deutsch*, soll in dieser Arbeit eine Überprüfung der Klassenebene erfolgen (siehe nächstes Kapitel, vgl. auch Kap. 7). Die hohen Intraklassenkorrelationen erklären sich dadurch, dass Fremdurteile, die auf Bekanntheit mit der einzuschätzenden Person aus derselben Situation basieren, einander stärker ähneln als Urteile von Personen, die die einzuschätzende Person aus unterschiedlichen Kontexten kennen (Asendorpf/Neyer 2012). Demzufolge ist es plausibel, dass die Lehrerpersönlichkeitsurteile der Schülerinnen und Schüler innerhalb einer Klasse stark miteinander korrelieren (vgl. Kap. 3.4.1.), was den guten Werten des dieser Ebene entsprechenden Reliabilitätsmaßes der Interklassenkorrelation (ICC 2) zwischen 0,87 und 0,93 entspricht ($>0,70$ gilt als befriedigend (vgl. Klusmann et. al. 2006, Klein et al. 2000, Stapleton et al. 2016)).

11.3.4 Mehrebenenmodelle der fachspezifischen Gesamtkonstrukte

Die vergleichsweise hohen Werte der Intraklassenkorrelationen der Skalen¹¹² sowie die guten Interklassenkorrelationen (ICC 2) legen eine Überprüfung der Big-Five-Struktur auf Klassenebene anhand der Erweiterung des Level-1-Modells um die Klassenebene (Level 2) nahe. Innerhalb dieser Mehrebenenanalysen wurde zugunsten des Gesamtkonstrukts *Persönlichkeit* darauf verzichtet, die jeweiligen Dimensionen separat zu berichten, obgleich diese – wie wohl angesichts der Komplexitätsreduktion kaum anders zu erwarten – ausgesprochen gute Model-Fit-Werte aufweisen, die deutlich robuster sind als jene, die hier verwendet werden. Das Vorgehen der simplifizierten bzw. separierenden Datenanalyse, das sich auf Skalen statt auf Konstrukte konzentriert, vernachlässigt jedoch die angemessene Interpretation der Modelle in ihrer Komplexität zugunsten passender Werte (vgl. Lang/Lüdtke 2005).

¹¹² Zwar lässt sich keine eindeutige Regel in der Fachliteratur finden, ab welchen Intraklassenkorrelationen Mehrebenenanalysen erfolgen sollten, jedoch weisen einige Autoren darauf hin, dass bereits geringe ICC-Werte eine Aussagekraft auf Klassenebene erwarten lassen (vgl. Kap. 7).

Es wurde die gekürzte Struktur¹¹³ des BFI-S als fachspezifisches Gesamtmodell (siehe hierzu Kap. 10, Abb. 8-10) in Deutsch und Mathematik sowohl auf Individual- (Modell 8) als auch auf Individual- und Klassenebene modelliert (Modell 9) und einem Modell gegenübergestellt, das auf Klassenebene nur noch einen einzigen Generalfaktor berücksichtigt (Modell 10). Anders als in den vorangegangenen Analysen wird hier also das Gesamtmodell *Persönlichkeit* betrachtet und nicht mehr dessen einzelne Dimensionen:

Tab. 13: (Mehrebenen-)Messmodelle im Vergleich (die Resilienzskala wurde um ein Item gekürzt; $N_{within} = 2239-2242/N_{between} = 123$).

Messmodell	Modellstruktur	Ebene	Chi ² /df	CFI	TLI	RMSEA	SRMR _w /SRMR _b	AIC/BIC
Deutsch								
4	BigFive	1	762,00/ 67	,93	,90	,07	,05	106464,07/ 106761,19
5	BigFive	1 & 2	1142,91/ 139	,89	,86	,06	,05/,12	104662,35/ 105148,02
6	BigFive + Generalfaktor	1 & 2	1928,06/ 149	,81	,76	,07	,06/,24	105153,87/ 105582,40
Mathematik								
4	BigFive	1	713,57/ 67	,94	,92	,07	,06	109092,17/ 109389,36
5	BigFive	1 & 2	1058,18/ 139	,90	,86	,05	,05/,12	106518,76/ 107004,55
6	BigFive + Generalfaktor	1 & 2	2278,55/ 149	,76	,71	,08	,06/,23	107291,41/ 107720,04

¹¹³ Gekürzt bedeutet: unter Ausschluss des anhand der vorausgegangenen Analysen ermittelten Items innerhalb der *Resilienz*-Skala.

Das Level 2-Generalfaktormodell dient der Überprüfung, ob sich die Faktorenstruktur der BigFive über die Individualebene hinaus auch auf der Klassenebene widerspiegelt oder ob auf der höheren Ebene (2) nur noch von einem allgemeinen, undifferenzierten Persönlichkeitsfaktor ausgegangen werden sollte. Die Fit-Werte untermauern jedoch die Annahme der Fünf-Faktoren-Struktur und legen deren Beibehaltung auch auf Klassenebene nahe. Grundlage für weitere Berechnungen ist demnach die Modellierung entlang des Messmodells 5, sprich: der Fünf-Faktoren-Struktur auf Individual- sowie auf Klassenebene (siehe Tabelle 13).

11.4 Messmodelle der Lehrerantworten

Das BigFive-Konstrukt innerhalb der Selbsteinschätzungen der eigenen Persönlichkeit durch die Lehrerinnen und Lehrer wurde ebenfalls zuerst mittels Reliabilitätsanalyse und im Weiteren mittels konfirmatorischer Faktorenanalyse überprüft. Im Unterschied zur Validierung der Schülereinschätzungen der Lehrerpersönlichkeit ist das Instrument zur Erhebung der Selbsteinschätzung bereits gut geprüft (Gerlitz/Schupp 2005) und kann daher direkt so übernommen werden. Zudem bedarf es innerhalb der Lehrerselbsteinschätzungen – entgegen dem Vorgehen bei den Schülereinschätzungen – keiner fachinternen Messmodelle, da die Lehrerinnen und Lehrer nicht unterschiedliche Lehrkräfte verschiedener Fächer, sondern lediglich sich selbst einschätzen mussten. Daher wurde ihre Persönlichkeitsselbsteinschätzung nicht fachspezifisch, sondern allgemein erfragt.

11.4.1 Reliabilität der einzelnen Skalen resp. Dimensionen¹¹⁴

Da sich innerhalb der Validierung der Schülerantworten eine fachspezifische Struktur abbildet (s.o.) und zur Beantwortung der ersten Forschungsfrage die Lehrerselbsteinschätzungen mit den Schülerfremdeinschätzungen der Lehrerpersönlichkeit in Zusammenhang gebracht werden, werden an dieser Stelle (Tab. 14) zwar zusätzlich die fachspezifischen Cronbachs α der Lehrerantworten berichtet, diese werden jedoch – aufgrund fehlender Argumentationsbasis – nicht weiter explizit auf ihre Fachspezifität analysiert. Allerdings lässt sich feststellen, dass die gekürzte Verträglichkeitsskala im Fach Deutsch im Prinzip nicht mehr als Skala gelten kann, wohingegen sie sich in Mathematik etwas stärker abbildet.

¹¹⁴ Gemäß dem Vorgehen der klassischen Testtheorie mittels Skalenanalyse per SPSS.

Berechnungen im Fach Deutsch, die auf Ergebnisse innerhalb der Dimension *Verträglichkeit* abzielen wurden daher nicht durchgeführt. Ein Überblick über die Passung der eingesetzten Skalen zu den erhobenen Lehrerdaten anhand der Cronbachs α -Werte stellt sich folgendermaßen dar:

Tab. 14: Cronbachs α der eingesetzten Skalen zur Erhebung der Lehrer-BigFive im Selbstbericht ($N=103$, davon $N_{\text{Deutsch}}=53$, $N_{\text{Mathematik}}=50$).

Dimension/ Skala	Cronbachs α (gekürzt)	
	fachübergreifend	Deutsch/Mathematik
Gewissenhaftigkeit	,66	,62/,71
Extraversion	,77	,67/,80
Offenheit für Erfahrungen	,70	,69/,56
Resilienz	,76	,73/,83
Verträglichkeit	,48 (,61 bzw. ,42)	,31/,55 (,05/,62)

Bis auf die Dimension *Verträglichkeit* weisen alle Dimensionen gute Skalenergebnisse auf. Zwar ließe sich diese Skala durch Auslassung des Items „*Ich bin ein Mensch, der verzeihen kann*“ auf ein Cronbachs α von 0,61 modifizieren, allerdings fordert Mplus eine andere Skalenstruktur ein: Modellberechnungen lassen die Verträglichkeitsskala nur dann zu, wenn das negativ gepolte Item „*Ich bin ein Mensch, der manchmal etwas grob zu anderen ist*“ ausgelassen wird. Dies verschlechtert den Cronbachs α zwar noch weiter, allerdings stellen die Berechnungen der Modelle mit Mplus die Skalenstruktur präziser heraus und auch hinsichtlich der inhaltlichen Validierung generell und im Besonderen bezogen auf Lehrerbefragungen fällt dieses Item als besonders schwierig zu beantworten bzw. als biasgefährdet – im Sinne der sozialen Erwünschtheit – auf: zwar bietet einerseits die auf intrapsychischen Informationen basierende selbstberichtete Persönlichkeitseinschätzung einen einmaligen und reflexiven Einblick in die Identität (Hogan/Roberts 2004, Roberts/Wood 2006), andererseits muss davon ausgegangen werden, dass diese Reflexion durch introspektive Begrenzungen (*blinde Flecken*, wie bspw. automatisierte Verhaltensweisen, die kaum reflektierbar und somit im Selbstbericht kaum direkt erfassbar sind (Back/Vazire 2012)) sowie starke Verzerrungsgefahr (bspw. durch Selbstverstärkung, -aufwertung und

-überschätzung, sozial erwünschtes Antwortverhalten und/oder Vergleichseffekte (Back/Vazire 2012, Srivastava 2012, Paulhus/Vazire 2007)) geprägt ist (vgl. Kap. 3.4.1). Vor diesem Hintergrund erscheint es plausibel, dass wohl kaum ein Lehrer oder eine Lehrerin als *grob* gegenüber den ihm oder ihr anvertrauten Schülerinnen und Schülern gelten möchte (oder sollte) und das Antwortverhalten demzufolge wenig reliabel ausfällt. Aufgrund dieser eher komplizierten Struktur der Dimension *Verträglichkeit* wird diese im weiteren Vorgehen lediglich in Form ihres Skalenmittelwerts entlang der Mplus-Lösung (also unter Auslassung des Items „...*manchmal etwas grob zu anderen...*“) innerhalb manifester Skalenmodelle berücksichtigt.

Während Mayr und Neuweg (2006) für die mittels NEO-PI-R erhobenen Persönlichkeitsdimensionen durchschnittliche Reliabilitäten von 0,87-0,91 bei den Lehrerinnen und Lehrern vorfinden, fallen die Reliabilitäten mit dem stark verkürzten Inventar des BFI-S hier – eher vergleichbar mit Befunden von Henoch et al. (2015), die von Reliabilitäten der Skalen zwischen 0,73 und 0,84 berichten – deutlich schwächer aus (Erklärungsansatz siehe Kap. 17.1).

11.4.2 Gesamtkonstrukt

Ogleich die Model-Fit-Werte entlang der etablierten Fünf-Faktoren-Struktur (Modell 2 in Tab. 15) alles andere als überzeugend sind, wird das Modell lediglich leicht modifiziert: Nach eingehender inhaltlicher Prüfung wurde das Item „*Ich bin ein Mensch, der manchmal etwas grob zu anderen ist*“, welches ein negativ gedrehtes Item zur Dimension *Verträglichkeit* darstellt, aus der Skala entfernt. Da plausibel erscheint, dass die Itemformulierung innerhalb eines implizit hierarchischen Lehrer-Schüler-Verhältnisses¹¹⁵ in besonderem Maße verstärkt zu einem sozial erwünschten Antwortverhalten seitens der Lehrkräfte einzuladen vermag oder dieses geradezu verlangt, wird darin der Grund dafür vermutet, dass das Item empirisch aus der Skala herausfällt (Faktorladungen sind dem Anhang C zu entnehmen).

¹¹⁵ Eine vertiefte Auseinandersetzung mit der hierarchischen Struktur von Lehrer-Schüler-Beziehungen findet sich bei Sternfeld (2010).

Tab. 15: Messmodelle der Lehrerantworten zu den erhobenen Persönlichkeitsdimensionen (N=123).

Messmodell	Modellstruktur	Chi ² / df	CFI	TLI	RMSEA	SRMR	AIC/ BIC
1	Generalfaktor	350,06/ 90	,44	,34	,17	,14	5172,84/ 5291,40
2	BFI-S	172,49/ 80	,80	,74	,11	,10	5015,27/ 5160,18
3	gekürzter BFI-S	138,99/ 67	,83	,76	,10	,08	4661,85/ 4798,86

Anmerkung: Messmodell 3 bildet den um ein Item in der Dimension *Verträglichkeit* gekürzten BFI-S ab.

Die Fit-Werte ändern sich durch die Kürzung (Modell 3) nur marginal und bleiben weiterhin sehr schwach ausgeprägt. Dies ist jedoch – wie bereits oben erwähnt – der komplexen Heterogenität der Facetten innerhalb der Dimensionen, der geringen Itemanzahl zur Erfassung sowie im Fall der Lehrerangaben der verhältnismäßig kleinen Stichprobe (N=103) geschuldet und somit in Kauf zu nehmen.¹¹⁶ Darüber hinaus weisen vergleichbare Inventare ebenfalls eher niedrige Model-Fit-Werte auf (siehe bspw. Rammstedt/Koch/Borg/Reitz 2004), möglicherweise wäre die Validität zu steigern, indem der berufliche Kontext direkt in die Itemformulierungen einbezogen wird (vgl. Hülshager/Specht/Spinath 2006).

¹¹⁶ Die besseren Fit-Indices bei den Schülerinnen und Schülern erklären sich zudem womöglich durch deren Aggregation auf Klassenebene, wodurch Beurteilungsfehler oder -tendenzen eher ausgeglichen werden.

12 Unterrichtsqualität: Instrument

Es wurden sieben Merkmale des Gesamtkonstrukts *Unterrichtsqualität* erfasst, um eine möglichst differenzierte Beschreibung der Unterrichtswahrnehmung der Schülerinnen und Schüler zu erhalten. Die Zusammenstellung der Merkmale und ihrer Skalen orientiert sich unter anderem am theoretischen Konstrukt guten Unterrichts nach Helmke (2009). Nach einer Beschreibung der in dieser Arbeit verwendeten Skalen zur Operationalisierung der Prozessmerkmale des Unterrichts erfolgt die Validierung des entsprechenden Antwortverhaltens der Schülerinnen und Schüler. Dabei wird entsprechend des obigen Vorgehens zunächst mittels konfirmatorischer Faktorenanalysen die Fachspezifität sowie Mehrebenenstruktur ermittelt und daraufhin die Testgüte der Skalen anhand klassischer Reliabilitätsanalysen überprüft.

Es wurden für die Untersuchungen der vorliegenden Arbeit vorab ausschließlich Aussagen zur Unterrichtsqualität ausgewählt, die in ihrer Formulierung eindeutig auf die subjektiven Einschätzungen der Schülerinnen und Schüler von unterrichtsgestaltenden Aktivitäten der Lehrkraft in Bezug auf die gesamte Klasse abzielen. Daher kommen hier Merkmale, die stark auf individualisierende oder differenzierende Unterrichtsprozesse fokussieren, nicht vor.

12.1 Prozessmerkmale der Unterrichtsqualität

Die in der Studie verwendeten Skalen zur Erfassung der schülerperzipierten Unterrichtsqualität entlang sogenannter Prozessmerkmale¹¹⁷ des Unterrichts sind teils leicht modifizierte oder gekürzte Fassungen von in der Unterrichtsforschung durchweg gut erprobten Instrumenten. Das eingesetzte Inventar präsentiert eine auf Prozessmerkmale des Unterrichts fokussierte Zusammenstellung aus folgenden sich als reliabel erwiesenen Untersuchungen zu Unterrichtsqualität:

- CES (Classroom Environment Scale von Moos und Trickett, 1974)
- Skalen zu Erziehungsumwelten (von Fend und Specht, 1986)
- LASSO (Landauer Skalen zum Sozialklima von Saldern und Littich, 1987)
- Scholastik-Projekt (von Helmke und Schrader bzw. Weinert, 1997)

¹¹⁷ Näheres dazu im entsprechenden Theorieteil (Kap. 4.2, 4.3).

- BIJU (Bildungsverläufe und psychosoziale Entwicklung im Jugendalter von Baumert/Gruehn et al. 1996)
- TIMSS (Third International Mathematics and Science Study von Beaton et al., 1996, bzw. Dritte internationale Mathematik- und Naturwissenschaftsstudie von Baumert/Lehmann et al. 1997)¹¹⁸

Folgende Konstrukte an Unterrichtsmerkmalen wurden für die Fragestellungen dieser Arbeit als relevant erachtet und ausgewählt, um die Unterrichtsqualität aus Schülersicht differenziert abzuzeichnen:

- *Angemessenes Interaktionstempo*
- *Angemessene Methodenvariation*
- *Zeitnutzung/Störungsfreiheit (Disziplin)*
- *Regelklarheit/Monitoring (Störungsprävention)*
- *Klarheit/Strukturiertheit*
- *Motivierung*
- *Alltagsbezug*

Die Antwortformulierungen von fünf der erhobenen Prozessmerkmale (*Angemessenes Interaktionstempo*, *Störungsprävention*, *Erklärkompetenz*, *Motivierung*, *Alltagsbezug*) beziehen sich direkt auf die Lehrkraft, während zwei der Skalen (*Angemessene Methodenvariation*, *Disziplin*) in ihrer Formulierung den Unterricht und demnach nur indirekt die Lehrkraft in den Mittelpunkt stellen.

Alle Skalen wurden anhand einer vierstufigen Likert-Skalierung (1=trifft nicht zu, 2=trifft eher nicht zu, 3=trifft eher zu, 4=trifft zu bzw. 1=(fast) immer, 2=oft, 3=manchmal, 4=nie) erfragt.

¹¹⁸ Der deutsche Teil von TIMSS (Third International Mathematics and Science Study von Beaton et al., 1996 bzw. Dritte internationale Mathematik- und Naturwissenschaftsstudie von Baumert/Lehmann et al. 1997) wurde durch Skalen zur Erfassung von Emotion und Motivation sowie zur Wahrnehmung von Unterricht erweitert, die jedoch auch bereits in der BIJU-Studie eingesetzt wurden. Es kommt daher in Bezug auf die Instrumente zu Überlappungen zwischen den verschiedenen Studien. In Bezug auf TIMSS wird im weiteren Verlauf lediglich die deutsche Quelle zitiert, da diese als Grundlage der Skalenbildung dient.

12.2 Beschreibung der einzelnen Skalen resp. Prozessmerkmale des Unterrichts

Die einzelnen Skalen des in dieser Studie eingesetzten Inventars zur Messung der Unterrichtsqualität anhand von Prozessmerkmalen werden im Folgenden zunächst in einem tabellarischen Überblick entlang ihrer Herkunft bzw. Entwicklung und ihres (ursprünglichen) Einsatzes dargestellt und anschließend semantisch analysiert.¹¹⁹

Im Folgenden wird ein Überblick über die eingesetzten Skalen gegeben, in deren Anschluss die jeweils erhobenen Prozessmerkmale der Unterrichtsqualität beschrieben und erläutert werden.

Tab. 16: *Eingesetzte Skalen zur Erfassung der Prozessqualität des Unterrichts – die kursiv gesetzten Items sind negativ zur jeweiligen Skala gepolt.*

Prozessmerkmal (Itemanzahl)	Beispiel-Item	Quelle(n)
Angemessenes Interaktions-tempo (3)	<i>Unsere Lehrerin/ unser Lehrer verlangt oft blitzschnelle Antworten.</i>	Baumert/Gruehn et al. (1996)
Angemessene Methoden-variation (3)	Dieser Unterricht ist methodisch abwechslungsreich.	Helmke/Weinert (1997)
Disziplin (4)	<i>In diesem Unterricht wird oft sehr gestört.</i>	Eder (1985), Fend/Specht (1986), Baumert/Gruehn et al. (1996), Baumert/Lehmann et al. (1997)
Störungsprävention (4)	Unsere Lehrerin/ unser Lehrer achtet sehr darauf, dass wir aufpassen.	Moos/Tricket (1974), Fend/Specht (1986), Baumert/Gruehn et al. (1996)

¹¹⁹ Zur Skalenbezeichnung negativ formulierte und im weiteren Verlauf der Analysen recodierte Items werden kursiv gedruckt.

Prozessmerkmal (Itemanzahl)	Beispiel-Item	Quelle(n)
Erklär- kompetenz (4)	Unsere Lehrerin/ unser Lehrer kann gut erklären.	Baumert/Gruehn et al. (1996), Saldern/Littig (1987)
Motivierung (3)	Unsere Lehrerin/ unser Lehrer kann auch trockenen Stoff wirklich interessant machen.	Baumert/Gruehn et al. (1996)
Alltagsbezug (3)	Um uns etwas zu erklären, nimmt unsere Lehrerin/ unser Lehrer oft ein Beispiel aus dem täglichen Leben.	Waldis et al. (2002)

12.2.1 Angemessenes Interaktionstempo

Das angemessene Interaktionstempo kann neben anderen Merkmalen der Kategorie Klassenführung und -organisation zugeordnet werden (vgl. Kastens 2009). Das Prozessmerkmal Interaktionstempo indiktorisiert eine quantitativ effektive Nutzung der Unterrichtszeit, die „[...] ein zügiges, aber dem Lernprozess der Schülerinnen und Schüler angepasstes Voranschreiten im Unterrichtsstoff (Interaktionstempo) [beinhaltet]“ (Kastens 2009, S. 44). Die Quantität der Unterrichtszeit gilt gleichauf mit deren Qualität als eine wichtige Bedingung – wenn nicht gar als die wichtigste (Baumert/Lehmann/Lehrke/Schmitz/Clausen/Hosenfeld/Köller/Neubrand 1997, S.199) – schulischer Lernprozesse (vgl. Carroll 1963) und ist zudem für die Verknüpfung von bestehendem und neu gewonnenem Wissen relevant (Helmke/Schrader 2008). Darüber hinaus ist die Angemessenheit des Unterrichtstempos Indikator für eine Passung zwischen Unter- und Überforderung im Unterricht (vgl. Helmke 2007).

Die Skala zur Erhebung der Angemessenheit des Interaktionstempos entspricht bis auf Kürzung eines Items¹²⁰ wortgetreu der Skala „Interaktionstempo“, die im Projekt BIJU (Baumert/Gruehn et al. 1996) entwickelt und erfolgreich validiert wurde. Ausgangspunkt ist die Annahme, dass in einem von den Schülerinnen

¹²⁰ Das Item „Unser Lehrer fragt oft unberechenbar in die Klasse und erwartet sofortige Antwort“ wurde ausgelassen.

und Schülern positiv wahrgenommenen Unterricht die Lehrkraft nicht unberechenbare und überhöhte Ansprüche an die unmittelbare Beantwortung direkt instruierter Fragen stellt, sondern sich durch eine verlässlich angemessene Rücksicht resp. Geduld auszeichnet. Insofern stellen die Items der Skala allesamt negativ gepolte Formulierungen der eigentlichen Skala dar.

12.2.2 Angemessene Methodenvariation

Angemessene Methodenvariation ist insofern als Qualitätsmerkmal anzusehen, als

„kein Lehrverfahren [...] für alle und für alles gleichermaßen geeignet [ist]. Sachgemäßer Methodenpluralismus und flexibles pädagogisches Handeln kennzeichnen deshalb guten Unterricht.“ (Helmke/Schrader 2008, S. 21)

Ein adäquater Einsatz verschiedener Methoden im Unterricht ist demzufolge den Lernprozessen der Schülerinnen und Schüler förderlich, indem interindividuell unterschiedliche Lernvoraussetzungen kompensiert werden können (vgl. Helmke/Schrader 2008, Kastens 2009, Gräsel/Göbel 2015). Dadurch wird die Möglichkeit einer Adaptivität des Unterrichts erhöht (vgl. Weinert 1997). Die Angemessenheit bezieht sich hier auf den Aspekt, dass ebenso einseitiger Frontalunterricht wie überbordende Methodenvielfalt der Unterrichtsqualität abträglich sind. Es geht dabei demnach nicht um ein Maximum an Methodenvariation, sondern um ein je nach Fach, Kontext und Lernziel abgewogenes Optimum (Helmke 2007).

In Anlehnung an das Scholastik-Projekt (Helmke/Schrader 1998) wurde die Skala zur Erhebung der *Angemessenheit der Methodenvariation* entwickelt. Diese Skala fragt die Methodenvielfalt im Unterricht ab. Die zwei positiv gepolten Items zielen auf eine Angemessenheit, nicht direkt auf die Häufigkeit der Wechsel von Methoden ab, da – wie Helmke (2014) deutlich macht – bei der Methodenvariation nicht zwangsläufig mehr gleich besser ist, sondern eher auch Zuviel des Guten sein kann, vergleichbar etwa mit einem „Finger in der Steckdose“ (Helmke 2014, S. 114). Ein negativ gedrehtes Item erfragt zudem die methodische Eintönigkeit.

12.2.3 Disziplin

Disziplin stellt einen wesentlichen Aspekt der übergeordneten Kategorie der effizienten Klassenführung dar und fördert ebenso wie das Interaktionstempo die effektive Zeitnutzung im Unterricht.

„Effiziente Klassenführung zielt durch Planung und vorbeugende Maßnahmen einerseits, durch situationsgemäße Flexibilität andererseits darauf ab,

Disziplinprobleme von vornherein zu vermeiden und ihnen gegebenenfalls angemessen zu begegnen.“ (Helmke/Schrader 2008, S. 21)

Für die eingesetzte Skala *Disziplin* wurden ein Item der Skala „Zeitverschwendung“¹²¹ entnommen, die aus den Skalen zu Erziehungsumwelten (Fend/Specht 1986) stammen, und drei aus der Skala „Unterrichtsstörungen“¹²², die im Rahmen des BIJU-Projekts Baumert/Gruehn et al. 1996) sowie TIMSS (Baumert/Lehmann et al. 1997) verwendet und dafür in Anlehnung an Eder (1985), Fend (1977), Fend & Specht (1986) entwickelt wurden. Diese Skala fragt die Effizienz der Zeitnutzung im Unterricht ab, indem sie die disziplinierte und unmittelbare Mitarbeit im Unterricht in den Blick nimmt. Alle vier Items stellen negativ gepolte Formulierungen zur Skala dar und beziehen sich indirekt auf verschiedene Arbeitsphasen im Unterricht, die bspw. durch Methoden- oder Sozialformwechsel zustande kommen (vgl. Gruehn 2000).

12.2.4 Störungsprävention

Störungsprävention zielt ebenfalls auf die Quantität der Unterrichtszeit resp. deren effektiven Nutzung ab und ist nach Kastens (2009) sowie Gräsel und Göbel (2015) ebenfalls unter die Kategorie Klassenführung und -organisation einzuordnen.

„Ein in diesem Sinne gelungener Unterricht zeichnet sich [weiter] dadurch aus, dass den Schülerinnen und Schülern die Regeln für den Unterricht deutlich sind (Regelklarheit) und die Lehrkräfte über effektive Mittel zum Umgang mit und in der Prävention von Unterrichtsstörungen verfügen.“ (Kastens 2009, S. 44)

Neben der Regelklarheit gilt das Monitoring im Sinne der Überwachung der Schülertätigkeiten als störungspräventive Maßnahme, die der effektiven Nutzung der Unterrichtszeit (und in dem Zuge dem Lernprozess der Schülerinnen

¹²¹ „In diesem Unterricht wird oft sehr gestört.“ „In diesem Unterricht wird fortwährend laut gequatscht.“

¹²² „In diesem Unterricht dauert es meistens sehr lange, bis alle zur Arbeit bereit sind.“ „In diesem Unterricht fehlt meistens bei irgend jemandem etwas, wenn wir anfangen wollen zu arbeiten.“

und Schüler) zugutekommt (Fend/Specht 1986). Außerdem kommt die Sicherung der Aufmerksamkeit der Schülerinnen und Schüler einer vertieften Verarbeitung zugute (vgl. Gräsel/Göbel 2015, siehe auch Helmke 2007).

Die Skala *Störungsprävention* setzt sich zusammen aus der Skala „Regelklarheit“¹²³ im CES (Moos und Tricket 1974) und einer leicht modifizierten Form der Skala „Monitoring“¹²⁴ aus BIJU (vgl. Gruehn 2000), die ihrerseits wiederum eine modifizierte und gekürzte Version der Skala „Störungspräventive Überwachung der Schülertätigkeiten“ aus den Skalen zu Erziehungsumwelten (Fend und Specht, 1986) darstellt.¹²⁵ *Regelklarheit* sowie *Überwachung der Schülertätigkeiten* bzw. der Einhaltung von Regeln dienen vor allem der quantitativ effektiven Zeitnutzung im Unterricht und sind als Aspekt der Klassenführung im Sinne Kounins (1976, vgl. außerdem Gruehn 2000, S. 129) zu verstehen. Demzufolge ist es als ein Merkmal guten Unterrichts anzusehen, wenn die Lehrkraft über die Vorgänge in der Klasse wacht und das Miteinander unmissverständlich so regelt, dass es bestenfalls erst gar nicht zu Störungen im Unterrichtsverlauf kommt.

12.2.5 Erklärkompetenz

Die Erklärkompetenz fokussiert vor allem die Klarheit und Strukturiertheit dargebotener Unterrichtsinhalte durch die Lehrkraft. Dabei gilt neues Wissen gezielt und altersgemäß mit bereits vorhandenem zu verknüpfen: „Wenn Wissen vernetzt werden soll, müssen Unterrichtsinhalte zudem sprachlich klar und gut

¹²³ „In diesem Unterricht hat unserer Lehrerin/ unser Lehrer klargemacht, was passiert, wenn man (Verhaltens-) Regeln verletzt.“

„In diesem Unterricht wird klar, was man machen darf und was nicht.“

¹²⁴ „Unsere Lehrerin/ unser Lehrer merkt sofort, wenn Schüler/innen beginnen, sich mit etwas anderem zu beschäftigen.“ Ursprüngliche BIJU-Formulierung: „Unsere Lehrerin/ unser Lehrer merkt sofort, wenn Schüler/innen beginnen, etwas anderes zu treiben.“

„Unsere Lehrerin/ unser Lehrer achtet sehr darauf, dass wir aufpassen.“

Ausgelassenes Item aus der BIJU-Skala: „Unsere Lehrerin/ unser Lehrer weiß immer genau, was in der Klasse vor sich geht.“

¹²⁵ Die Formulierung des ersten Items lautete ursprünglich: „Unsere Lehrerin/ unser Lehrer merkt sofort, wenn Schüler/innen beginnen, etwas anderes zu treiben“, anstatt „Unsere Lehrerin/ unser Lehrer merkt sofort, wenn Schüler/innen beginnen, sich mit etwas anderem zu beschäftigen“ wie es im Rahmen dieser Studie formuliert wurde.

strukturiert [...] dargeboten werden.“ (Helmke/Schrader 2008, S. 21, siehe hierzu auch Gräsel/Göbel 2015). Darüber hinaus

„[zeichnet sich] ein klar strukturierter Unterricht [...] dadurch aus, dass die Lehrkraft regelmäßig die neu diskutierten Inhalte zusammenfasst, wiederholt und mit bereits Bekanntem vernetzt (Clausen, 2002; Klieme, 2002b; Kunter et al., 2005; Seidel, Rimmele, & Prenzel, 2005). Dabei sollten den Schülerinnen und Schülern zu jedem Zeitpunkt die Unterrichtsziele und Aufgabenstellungen klar und die einzelnen Unterrichtsschritte nachvollziehbar sein (Brophy, 1999; Seidel, Prenzel, Duit, Euler, Geiser, & Hoffmann et al., 2002; Trepke, 2004).“ (Kastens 2009, S. 46)

Die eingesetzte Skala fasst die beiden Subskalen „Strukturiertheit des Unterrichts,“ und „Klarheit der Präsentation“ der BIJU-Skala „Klarheit und Strukturiertheit des Unterrichts“ (Baumert/Gruehn et al. 1996) zusammen, die angelehnt an LASSO (Saldern/Littich 1987) entwickelt wurden¹²⁶. Die Skalen wurden für die vorliegende Studie gekürzt und leicht modifiziert.

Die verwendeten Items fragen die Verständlichkeit des im lehrergeleiteten Unterricht für den Lernerfolg nicht unwesentlichen Lehrervortrags ab. Dabei sind vor allem Strukturierungshilfen bzgl. des dargebotenen Lernstoffs bedeutsam.

12.2.6 Motivierung

Motivierung ist für das schulische Lernen insofern relevant, als dass sie auf die nachweislich lernförderliche intrinsische Motivation abzielt. Durch das strategische Ansprechen von Neugier und Interesse, ergänzt um Aspekte extrinsischer Motivierung wie positiver Verstärkung resp. Lob sollen Lerninhalte besser vernetzt (vgl. Deci/Ryan 1993, Helmke/Schrader 2008). Neben der thematischen Motivierung – durch Auswahl und Gestaltung von Lerninhalten – sowie instrumentellen Motivierung – durch Betonung von Nützlichkeit und Bedeutsamkeit der Lerninhalte – gilt die mitreißende Begeisterung der Lehrkraft hierbei als fundamental (vgl. Helmke 2007).

¹²⁶ Das Item „Unsere Lehrerin/ unser Lehrer fasst häufig nochmals den Stoff zusammen, damit wir ihn uns gut merken können“ wurde bereits identisch in LASSO (Saldern/Littich 1987) verwendet.

Die *Motivierungsskala* ist identisch zur Skala „Motivierungsfähigkeit“ (oder auch „Motivierung durch den Lehrer“ genannt), die in BIJU (Baumert/Gruehn et al. 1996) entwickelt wurde. Die Items sind in Anlehnung an Krapps Interessentheorie (1992) formuliert und beziehen sich auf das unterrichtliche Ziel im Unterrichtsprozess Interesse bzw. Interessen zu wecken¹²⁷. Grundlage dafür dieses Merkmal als Indikator guten Unterrichts aufzuführen ist die Erkenntnis, dass das Interesse an einem Gegenstand – auch extrinsisch indiziertes – die Lernmotivation bezüglich dieses Gegenstands steigert und damit auch die Vernetzung von Lerninhalten fördert (vgl. Deci/Ryan 1993).

12.2.7 Alltagsbezug

Der Alltagsbezug impliziert Ansätze konstruktivistischer Lerntheorien und kann in die übergeordneten Qualitätskategorien der konstruktivistischen Orientierung oder der kognitiven Aktivierung zugeordnet werden. Basis konstruktivistischer Lerntheorien ist die Annahme, dass Lernen ausschließlich in Form eines selbstständigen, d.h. selbstbestimmten und selbstgesteuerten Prozesses gelingen kann. Voraussetzungsvoll für den Lernprozess ist dabei die Verknüpfung neuen Wissens mit bereits bestehendem, wodurch Erkenntnisse erneuert bzw. erweitert werden (vgl. Kastens 2009).

„Ein solcher Unterricht zeichnet sich unter anderem dadurch aus, dass das Lernen neuer Inhalte unter Berücksichtigung des Vorwissens der Schülerinnen und Schüler stattfindet und Inhalte kontextgebunden (Alltagserfahrungen, Inhalte anderer Unterrichtsfächer) vermittelt werden.“ (Kastens 2009, S. 45)

Darüber hinaus stellt die Einbindung von Alltagserfahrungen der Schülerinnen und Schüler eine Möglichkeit dar, dass diese sich kognitiv aktiv beteiligen, indem sie ihre eigenen Erfahrungen hinterfragen und auf andere Kontexte übertragen (vgl. Kastens 2009).

Die entsprechende Skala ist eine modifizierte, gekürzte Form der Skala „Motivierung“ von Waldis et al. (2002; zit. n. Rakoczy/Buff/Lipowsky 2005, S. 57). Der Skala liegt die Annahme zugrunde, dass die Einbindung der alltäglichen Umwelt von Schülerinnen und Schülern deren Interesse für die unterrichtlichen

¹²⁷ Interesse in diesem Sinne verstanden gilt als *State* und steht mit seinem *situationalen* Charakter im Gegensatz zu einem bereits bestehendem *individuellen* Interesse, das als *Trait* bezeichnet wird (vgl. Krapp 2014).

Lerngegenstände verstärkt weckt und sie dadurch verstärkt für den Lernprozess motiviert, was seinerseits wiederum zu besseren Lernerfolgen führt (vgl. Deci/Ryan 1993). Die Itemformulierungen zielen daher auf die Erfassung der Verknüpfung von Lerninhalten und dem Schüleralltag ab.

12.3 Validierung des Konstrukts der Unterrichtsqualität

Angesichts der Besonderheiten des Erhebungsdesigns und erster Ergebnisse der Hauptuntersuchung (siehe Rahn/Fuhrmann/Keune/Gruehn, im Druck) – zum einen bezüglich der Tatsache, dass Schülerinnen und Schüler ganzer Klassen erhoben wurden (Näheres hierzu siehe unter Kap. 7), zum anderen, dass diese nacheinander, wenn auch in unterschiedlicher Reihenfolge¹²⁸, zu mehreren Fächern¹²⁹ befragt wurden – scheint es geboten, die Skalen der Prozessmerkmale zur Erhebung der Unterrichtsqualität trotz ihres bereits etablierten Einsatzes in diversen Studien präzise hinsichtlich möglicherweise fachspezifischen Antwortverhaltens zu validieren. Dies erfolgt entlang verschiedener Messmodelle (siehe Vier-Modell-Prüfung in Kap. 10), die in unterschiedlicher Gewichtung fachspezifisches Antwortverhalten einbeziehen, um die Passung der Daten an die entsprechenden theoretischen Konstrukte und Hypothesen zu ermitteln.

Die verschiedenen Messmodelle werden im Folgenden aufeinander aufgebaut, um die verschiedenen Analyseschritte möglichst nachvollziehbar darzustellen. In einem ersten Verfahren wird mithilfe konfirmatorischer Faktorenanalysen die Fachspezifität des Antwortverhaltens herausgestellt. Erst danach werden vollständigshalber die Skalen zur Ermittlung der wahrgenommenen Unterrichtsqualität entlang der Prozessmerkmale des Unterrichts fachspezifisch mittels klassischer Reliabilitätsanalyse betrachtet (Cronbachs α , ICC 1+2). Abschließend erfolgt eine Analyse des Gesamtmodells *Prozessmerkmale der Unterrichtsqualität* unter Berücksichtigung der Mehrebenenstruktur der Daten.

¹²⁸ Dies erfolgt, um Reihenfolgebias vorzubeugen (vgl. Kap. 6.2).

¹²⁹ Die Schülerinnen und Schüler wurden alle zu den Fächern Deutsch und Mathematik und an den Berufskollegs zusätzlich zu einem profilbildenden Fach befragt (Näheres hierzu in Kap. 6.2).

12.3.1 Vier-Modell-Prüfung

Im weiteren Verfahren wurden die Antworten bezüglich der erhobenen Merkmale mittels konfirmatorischer Faktorenanalysen auf ihre fachinterne Validität getestet, indem ein fachübergreifendes, einfaktorielles Modell (siehe Tab. 17, Messmodell 1), ein fachspezifisch ausgelegtes Modell (siehe Tabelle 17 Messmodell 3a und b) sowie ein genestetes Modell (siehe Tabelle 17 Messmodell 2) anhand ihrer jeweiligen *Model-Fit*-Werte miteinander verglichen wurden. Im einfaktoriellen Messmodell 1 wird davon ausgegangen, dass keine Fachunterschiede im Antwortverhalten bestehen (vgl. Abb. 4, Kap. 10), während das genestete Modell (2) sowohl die unkorrelierte fachspezifische Ausrichtung des Antwortverhaltens als auch einen gemeinsamen fachübergreifenden Generalfaktor einbezieht (vgl. Abb. 5, Kap. 10). Dabei wird davon ausgegangen, dass die Schülerinnen und Schüler zwischen den beiden Fächern differenzieren, den Antworten jedoch gleichzeitig in beiden Fächern dasselbe Antwortverhalten zugrunde liegt. Das fachspezifische Modell (3a) dagegen geht davon aus, dass das Antwortverhalten gemäß der Fächertrennung zu betrachten ist, jedoch ein methodischer Faktor (durch den Einsatz identischer Instrumente) in Interfachkorrelationen sichtbar wird (vgl. Abb. 6, Kap. 10). In Messmodell 3b wird ausschließlich das fachspezifische Antwortverhalten ohne fachübergreifende Korrelationen analysiert (Abb. 7, Kap. 10). Diesem Modell zufolge differenzieren die Schülerinnen und Schüler vollständig zwischen den Fächern.

Unterrichtsqualität: Instrument

Tab. 17: Messmodelle der schülerperzipierten Prozessmerkmale des Unterrichts (N=2248-2250).

Prozess-merkmal	Messmodell	Chi ² /df	CFI	TLI	RMSEA	SRMR	AIC/ BIC
<i>Angemessene Methodenvariation</i>	1	2178,79/9	,51	,18	,33	,24	32311,44/ 32414,36
	2	5,46/3	1	1	,02	,02	29409,27/ 29546,50
	3a	13,93/8	1	1	,02	,01	29409,94/ 29518,57
	3b	16,37/9	1	1	,02	,02	29411,90/ 29514,82
<i>Motivierung</i>	1	10586,16/ 9	,25	-,26	,72	,27	31911,39/ 32014,32
	2	16,20/3	1	1	,04	,02	28283,36/ 28420,60
	3a	29,07/8	1	1	,03	,01	28290,43/ 28399,08
	3b	28,36/9	1	1	,03	,03	28292,38/ 28395,31
<i>Angemessenes Interaktionstempo</i>	1	1083,62/9	,64	,40	,23	,16	32966,37/ 33069,31
	2	17,75/3	1	,98	,05	,03	31883,21/ 32020,46
	3a	71,36/8	,98	,96	,06	,03	31909,04/ 32017,70
	3b	82,46/9	,98	,96	,06	,04	31923,24/ 32026,17
<i>Alltagsbezug</i>	1	4395,67/9	,31	-,16	,47	,254	33808,50/ 33911,43
	2	8,04/3	1	1	,03	,01	30638,70/ 30775,94
	3a	14,82/8	1	1	,02	,01	30636,93/ 30745,58
	3b	23,39/9	1	1	,03	,04	30647,02/ 30749,95

Unterrichtsqualität: Instrument

Prozess-merkmal	Mess-modell	Chi ² /df	CFI	TLI	RMSEA	SRMR	AIC/ BIC
Erklärkompetenz	1				n.i.		
	2	88,34/12	,99	1	,05	,02	42665,60/ 42848,59
	3a	164,90/19	,98	,97	,06	,04	42785,01/ 42927,97
	3b	158,80/20	,98	,98	,06	,04	42783,65/ 42920,89
Störungsprävention	1	4249,96/20	,43	,21	,31	,21	44960,80/ 45098,02
	2	59,78/12	,99	,99	,04	,03	41863,18/ 42046,14
	3a	255,67/19	,97	,95	,07	,04	42073,47/ 42216,42
	3b	257,21/20	,97	,96	,07	,05	42088,11/ 42225,34
Disziplin	1	4894,16/20	,37	,12	,33	,19	43930,50/ 44067,74
	2				n.i.		
	3a	163,08/19	,98	,97	,06	,04	41445,85/ 41588,80
	3b	179,82/20	,98	,97	,06	,07	41485,12/ 41622,36

Anmerkung: n.i. steht für nicht identifizierbar, was bedeutet, dass sich das Modell nicht anhand der vorgegebenen Parameter berechnen lässt.

Insgesamt lässt sich feststellen, dass ein Generalfaktormodell, das Fachunterschiede im Antwortverhalten ignoriert, ausgeschlossen werden kann, da dieses sowohl in der Theorie als auch auf der Basis der vorliegenden Daten deutliche Schwächen zeigt. Vielmehr ist davon auszugehen, dass das Antwortverhalten fachspezifisch erfolgt: Zwischen einem Fächermodell und dem genesteten Mo-

dell zeigen sich minimale Unterschiede, wonach entlang empirischer Argumentation das einfachere (d.h. fächergetrennte) Modell für weitere Analysen verwendet werden sollte. Zumal die Interfachkorrelationen sehr gering ausfallen:

Tab. 18: Deutsch-Mathematik-Korrelationen (manifest auf Individualebene) innerhalb der einander entsprechende Dimensionen (Signifikanz auf 0,01-Niveau).

Methoden-variation	Motivierung	Interaktionstempo	Alltagsbezug	Erklärkompetenz	Störungsprävention	Disziplin
,05	,05	,11**	,08**	-,02	,10**	,16**

12.3.2 Reliabilität der einzelnen Skalen

In der folgenden Tabelle sind die Reliabilitäten (Cronbachs α und ICC 1+2) der Skalen fachintern zusammengefasst:

Tab. 19: Reliabilität der Prozessmerkmale aus Schülersicht (N=2248-2250).

Prozessmerkmale (Itemanzahl)	Cronbachs α		ICC 1		ICC 2	
	Deutsch	Mathe	Deutsch	Mathe	Deutsch	Mathe
Ang. Interaktionstempo (3)	,68	,76	,14	,25	,76	,86
Ang. Methodenvariation (3)	,85	,83	,16	,17	,78	,80
Disziplin (4)	,79	,82	,29	,33	,88	,90
Störungsprävention (4)	,82	,81	,22	,25	,83	,86
Erklärkompetenz (4)	,79	,84	,25	,38	,86	,91
Motivierung (3)	,88	,90	,23	,36	,83	,91
Alltagsbezug (3)	,86	,86	,16	,26	,77	,86

Die Cronbachs α sind durchweg zufriedenstellend und die Skalen bedürfen keiner weiteren Prüfung oder Modifikation. Die Zweckmäßigkeit etablierter Skalen zur Unterrichtsqualität kann dadurch als erneut bestätigt erachtet werden.

Angesichts der eher hoch ausgeprägten ICC-Werte, die für eine hohe gemeinsame Varianz innerhalb der Klassen sprechen (ICC 1 von 0,14-0,38), scheint es allerdings durchaus angemessen zu prüfen, inwiefern die Klassenebene der Berücksichtigung bedarf (vgl. Kap. 7).

12.3.3 Mehrebenenmodelle der fachspezifischen Gesamtkonstrukte

Nach dieser Prüfung der fachinternen Validität werden fachspezifische Gesamtmodelle des Konstrukts *Unterrichtsqualität* jeweils für das Fach Deutsch und das Fach Mathematik – zunächst auf Individualebene (Modell 4, siehe Abb. 8, Kap. 10) – erstellt. Entlang dieses Gesamtkonstrukts erfolgen nun Mehrebenenanalysen unter Einbezug der Klassenebene (als Level-2-Einheit): Während Messmodell 5 die Faktorenstruktur gemäß den Prozessmerkmalen auch auf Klassenebene beibehält (siehe Abb. 9, Kap. 10), wird dieser in Messmodell 6 eine Generalfaktorenstruktur auf Klassenebene (siehe Abb. 10, Kap. 10) gegenübergestellt:

Tab. 20: modifizierte (Mehrebenen-)Messmodelle der Prozessmerkmale ($N_{within}=2248-2250$, $N_{between}=123$).

Messmodell	Modellstruktur	Ebene	Chi ² / df	CFI	TLI	RMSEA	SRMR _w / SRMR _b	AIC/ BIC
Deutsch								
4	Prozessmerkmale	1	954,76/ 231	,97	,96	,04	,04	116804,33/ 117336,04
5	Prozessmerkmale	1 & 2	1544,60/ 469	,95	,95	,03	,03/,10	115170,83/ 116057,02
6	Prozessmerkmale + Generalfaktor	1 & 2	2553,06/ 490	,91	,90	,04	,04/,34	115835,77/ 116601,89

Unterrichtsqualität: Instrument

Messmodell	Modellstruktur	Ebene	Chi ² /df	CFI	TLI	RMSEA	SRMR _w /SRMR _b	AIC/BIC
Mathematik								
4	Prozessmerkmale	1	995,38/ 231	,97	,96	,04	,04	120862,72/ 121394,47
5	Prozessmerkmale	1 & 2	1621,35 / 469	,95	,94	,03	,03/,09	118432,84/ 119319,10
6	Prozessmerkmale + Generalfaktor	1 & 2	2936,24 / 490	,89	,88	,05	,04/,33	119361,22/ 120127,40

Es zeigt sich, dass trotz deutlicher Verschlechterung des Chi²/df, leichter Verschlechterung des CFI- und des TLI-Index‘ die Werte des RMSEA sowie AIC und BIC in Messmodell 5 am besten sind. Da bei CFI, TLI und vor allem bei Chi²/df anzunehmen ist, dass das Absinken der Model-Fit-Werte stark durch die gesteigerte Modellkomplexität beeinflusst sind, wird hier den Fit-Werten des direkten Modellvergleichs, nämlich AIC und BIC gefolgt. Demnach kann das Modell 6 als das Model, welches die Daten am differenziertesten und zugleich am besten widerspiegelt, angesehen werden.

Insgesamt lässt sich in Bezug auf bestehende Befunde aus der Unterrichtsfor- schung festhalten, dass die hier befragten Schülerinnen und Schüler sehr deutlich zwischen den diversen Prozessmerkmalen differenzieren.¹³⁰

¹³⁰ Grundlage ist der Befund, dass je jünger befragte Schülerinnen und Schüler sind, ihre Angaben zur Unterrichtsqualität desto undifferenzierter erfolgen und sie nicht so wie Schülerinnen und Schüler älterer Jahrgänge – etwa der Jahrgangsstufe 11 wie in der vor- liegenden Untersuchung – zwischen verschiedenen Merkmalen oder Dimensionen des Unterrichts zu unterscheiden vermögen (vgl. Ditton 2002a).

13 Analysestrategien

Bevor in einem ersten Analyseschritt die deskriptive Auswertung der erfassten Konstrukte – und im Fall des Konstrukts *Lehrerpersönlichkeit* zudem die Auswertungen der jeweiligen Perspektiven – erfolgt (siehe Kap. 14), wird zunächst erläutert, welche Analysen zur Beantwortung der jeweiligen Forschungsfragen durchgeführt werden.

13.1 Forschungsfrage 1: Welcher Zusammenhang besteht zwischen selbst- und fremdwahrgenommener Lehrerpersönlichkeit?

Die vorab zu klärende und somit erste Forschungsfrage bezieht sich auf Übereinstimmungen zwischen den Einschätzungen der Lehrerpersönlichkeit durch die Schülerinnen und Schüler sowie deren Lehrkräfte selbst: Inwiefern stimmen Selbst- und Fremdeinschätzungen der Lehrerpersönlichkeit überein bzw. lassen sich Differenzen zwischen den verschiedenen Perspektiven beobachten?

Dazu werden zunächst Korrelationen zwischen den entsprechenden BigFive-Dimensionen der Persönlichkeitseinschätzungen durch die Schülerinnen und Schüler mit ihren Lehrerinnen und Lehrern berechnet und ausgewertet. Dies erfolgt zum einen auf Individualebene, wobei die Selbsteinschätzungen der jeweiligen Lehrkräfte den Einschätzungen durch ihre Schülerinnen und Schüler einzeln gegenübergestellt werden. Dadurch erfolgt im Datensatz eine künstliche Vervielfältigung der Lehrerangaben, da diese in diesem Fall als Merkmal der entsprechenden Schülerinnen und Schüler gehandhabt werden. Zum anderen werden die Korrelationen auf der Klassenebene berechnet. In diesem Fall wird der jeweiligen Lehrerselbsteinschätzung der auf Klassenebene aggregierte Wert der Schülereinschätzungen gegenübergestellt. Die beschriebenen Analyseschritte werden der Validierung der Schülerantworten, die deutlich auf ein fachspezifisches Antwortverhalten hinweisen, entsprechend fachspezifisch für die Fächer *Deutsch* und *Mathematik* durchgeführt. Zudem werden die Korrelationen auf manifester Skalenebene berechnet, da ansonsten zum einen die Modelle für Mplus zu komplex zu berechnen sind, und zum anderen in der Dimension *Resilienz* zwar Konstruktidentität besteht, jedoch durch die vorgenommene Kürzung keine identischen Skalen vorliegen. Die Dimension *Verträglichkeit* wird darüber hinaus

nicht in die Korrelationsanalysen einbezogen, da bei den Lehrkräften angesichts der Reliabilitätsmaße keine Skala vorausgesetzt werden kann.



Abb. 12: vereinfachtes Modell über Korrelationen zwischen den verschiedenen Lehrer-Big-Five-Dimensionen aus Lehrer- und Schülersicht.

Neben den bivariaten Korrelationen zwischen Lehrer- und Schülerangaben, die Zusammenhänge innerhalb der einzelnen Dimensionen abbilden, wird außerdem eine Korrelationsmatrix über alle Dimensionen erstellt, um mögliche Zusammenhänge über die jeweilige Dimension hinaus zu überprüfen.

13.2 Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten *Lehrerpersönlichkeit* und *Unterrichtsqualität*?

Zur Untersuchung der zweiten Forschungsfrage, die sich mit dem Zusammenhang der Konstrukte *Lehrerpersönlichkeit* und *Unterrichtsqualität* befasst, werden entsprechende Regressionsmodelle hier ebenfalls ausschließlich fachspezifisch für Deutsch und Mathematik erstellt und verglichen. Dies gebietet die Validierung der erhobenen Prozessmerkmale des Unterrichts durch die Schülerinnen und Schüler zur Ermittlung der wahrgenommenen Unterrichtsqualität, die auf fachspezifische Schülerfeedbacks hinweist (vgl. Kap. 12.3). Es wurde deutlich, dass die Schülerurteile zwischen den Fächern differenziert ausfallen und es demnach nicht legitim erscheint, diese Differenzen beispielsweise durch die Analyse fachübergreifender Modelle zu ignorieren. Zudem wurde bereits deutlich, dass latente Gesamtmodelle für Mplus zu komplex zu berechnen sind. Daher werden die Modelle auf Basis der manifesten Skalen berechnet.

Im ersten Teil der Fragestellung wird ermittelt, inwiefern ein perspektivenübergreifender Zusammenhang der Persönlichkeitselbsteinschätzungen der Lehrkräfte mit schülerperzipierten Prozessmerkmalen des Unterrichts besteht. Dazu wird zunächst eine Korrelationsmatrix der selbsteingeschätzten Lehrer-BigFive und den schülerperzipierten Prozessmerkmalen des Unterrichts erstellt, um abzuschätzen, inwieweit die Höhe der Korrelationen weitere Regressionsanalysen rechtfertigen.

Im zweiten Teil der Fragestellung, gleichwohl Hauptaugenmerk der vorliegenden Arbeit, wird perspektivenspezifisch und auf den berichteten Forschungsbefunden zur Lehrerpersönlichkeit (Kap. 2.2) aufbauend untersucht, welchen Einfluss die Lehrerpersönlichkeit auf die Unterrichtsqualität hat – beides aus Perspektive der Schülerinnen und Schüler und simultan auf Individual- und Klassenebene, um sowohl individuelle als auch Klasseneffekte zu untersuchen. Es handelt sich hierbei um einen explorativen Ansatz, da Prozessmerkmale des Unterrichts bislang nicht auf ausdifferenzierte, schülerperzipierte Lehrer-BigFive regrediert wurden. Daher lassen sich keine theoretisch abgeleiteten Hypothesen bezüglich der Effekte ableiten. Um einen differenzierten Überblick über die Zusammenhänge zu gewinnen, wird zuerst ein saturiertes multiples Regressionsmodell berechnet. Dazu werden manifeste multiple Strukturmodelle (in Deutsch und Mathematik) entlang einer Regression der sieben Prozessmerkmale des Unterrichts auf die fünf Persönlichkeitsdimensionen erstellt (Modell 1a). Daraufhin wird ein Modell geprüft, dass ausschließlich die ermittelten signifikanten Pfade berücksichtigt (1b).

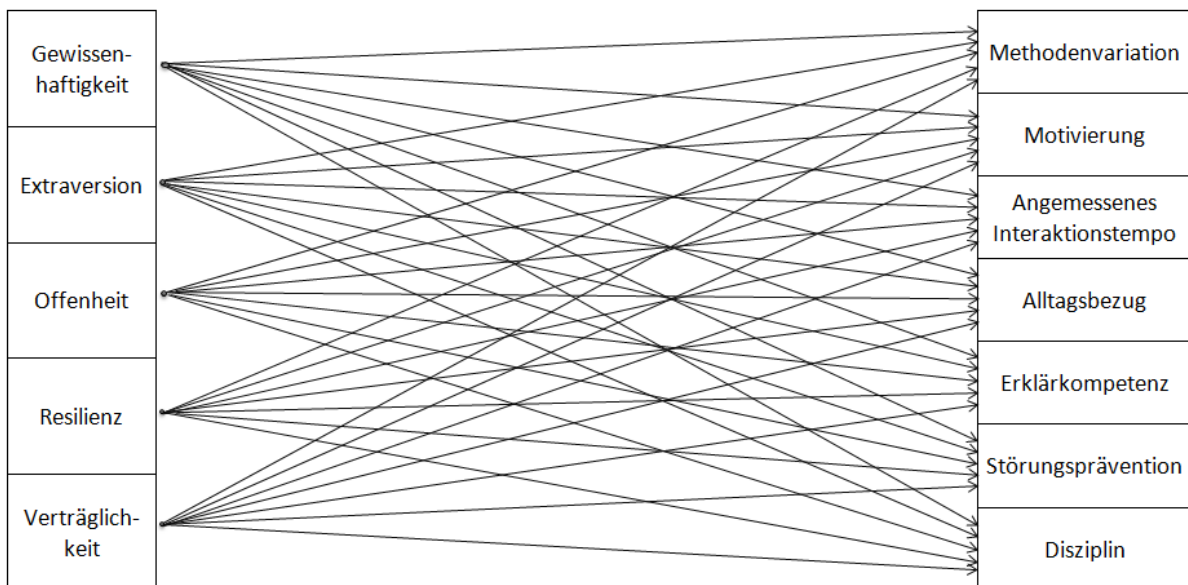


Abb. 13: saturiertes Strukturmodell mit Effekten der Lehrerpersönlichkeit auf die Unterrichtsqualität (Modell 1a).

Angesichts dessen, dass die Prädiktoren alle zusammen das Konstrukt *Persönlichkeit* abbilden, muss davon ausgegangen werden, dass diese miteinander korreliert sind bzw. sind Interkorrelationen der Persönlichkeitsdimensionen dem Konstrukt *Persönlichkeit* immanent (siehe hierzu Kap. 14.2.4). Es handelt sich

angesichts der inhaltlichen Zusammenhänge um eine essentielle Multikollinearität, die kaum zu bereinigen ist (vgl. Urban/Mayerl 2011).¹³¹ Mplus berücksichtigt bei Regressionsanalysen die Korrelationen zwischen den Prädiktoren per Voreinstellung und rechnet diese heraus (würden diese auf 0 programmiert, dann würden die Signifikanzen fälschlicherweise steigen). Demnach tangiert eine Multikollinearität die Signifikanzindikatoren (p-values) nicht (siehe Muthén & Muthén, 2016). Im Diskussionsforum von Mplus beantwortet Bengt Muthén, einer der Entwickler des Statistikprogramms, die Frage nach der Problematik mit Multikollinearität zudem wie folgt:

„I don't think this is problematic. Factors are often naturally correlated, not independent. Multicollinearity is always a risk. One way around this is to postulate a second order factor behind highly correlated factors. [...] I think multicollinearity can be checked just like in regular regression with observed variables. With a 2nd-order-factor, you can simply regress the DV on that factor only.” (Muthén 2017)

Angesichts der Verzerrungsgefahr der Regressionskoeffizienten sowie der Varianzaufklärung durch eine theoretisch anzunehmende und konstruktimmanente Multikollinearität der einzelnen Persönlichkeitsdimensionen, erscheint eine genauere Betrachtung der Zusammenhänge jedoch erforderlich (Näheres zu diesem Vorgehen siehe Stapleton et al. 2016, Urban/Mayerl 2011). Wooldridge (2012) sowie Allison (2012) schlagen wie Muthén (2017) vor, sehr hoch ($r > 0,9$) korrelierte Prädiktoren zu einem gemeinsamen Faktor 2. Ordnung zusammenzufassen. Da die Schülerinnen und Schüler die BigFive-Dimensionen ihrer Lehrkräfte jedoch zu differenzieren scheinen, da die Dimensionen in einer explorativen Faktorenanalyse getrennte Faktoren abbilden (siehe Kap. 11.3.1), werden möglicherweise bestehende Kollinearitäten zunächst in Form von modellierten Korrelationen (vgl. hierzu Kap. 14.2.4) in die Strukturmodelle aufgenommen (Modell 2a). Desweiteren wird auch hier entlang der Befunde der vorangegangenen Analysen ein Modell berechnet, das ausschließlich die signifikanten Pfade der verschiedenen Persönlichkeitsdimensionen (sowie ggf. deren relevanten Korrelationen) berücksichtigt (Modell 2b).

¹³¹ Diese Form steht im Gegensatz zur nicht-essentiellen Multikollinearität, deren Linearität gezielt konstruiert wird (vgl. Urban/Mayerl 2011).

Analysestrategien

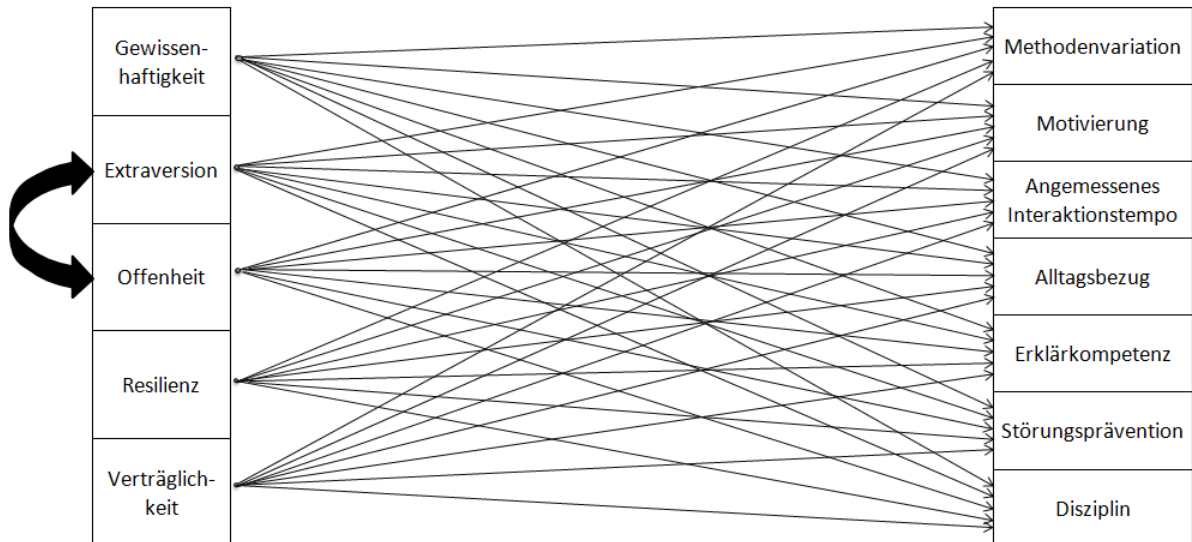


Abb. 14: Strukturmodell mit Effekten der Lehrerpersönlichkeit auf die Unterrichtsqualität mit Korrelationen zwischen sehr hoch korrelierenden Prädiktoren (Modell 2a).

Abschließend wird ein Faktor 2. Ordnung (im Sinne eines Gesamtkonstrukts *Persönlichkeit*) über alle BigFive-Dimensionen gebildet, um den Effekt des Gesamtkonstrukts auf die Prozessmerkmale der Unterrichtsqualität innerhalb einer linearen Regression abzubilden und das Ausmaß der Verzerrung durch die Multikollinearität abzuschätzen (Modell 3).

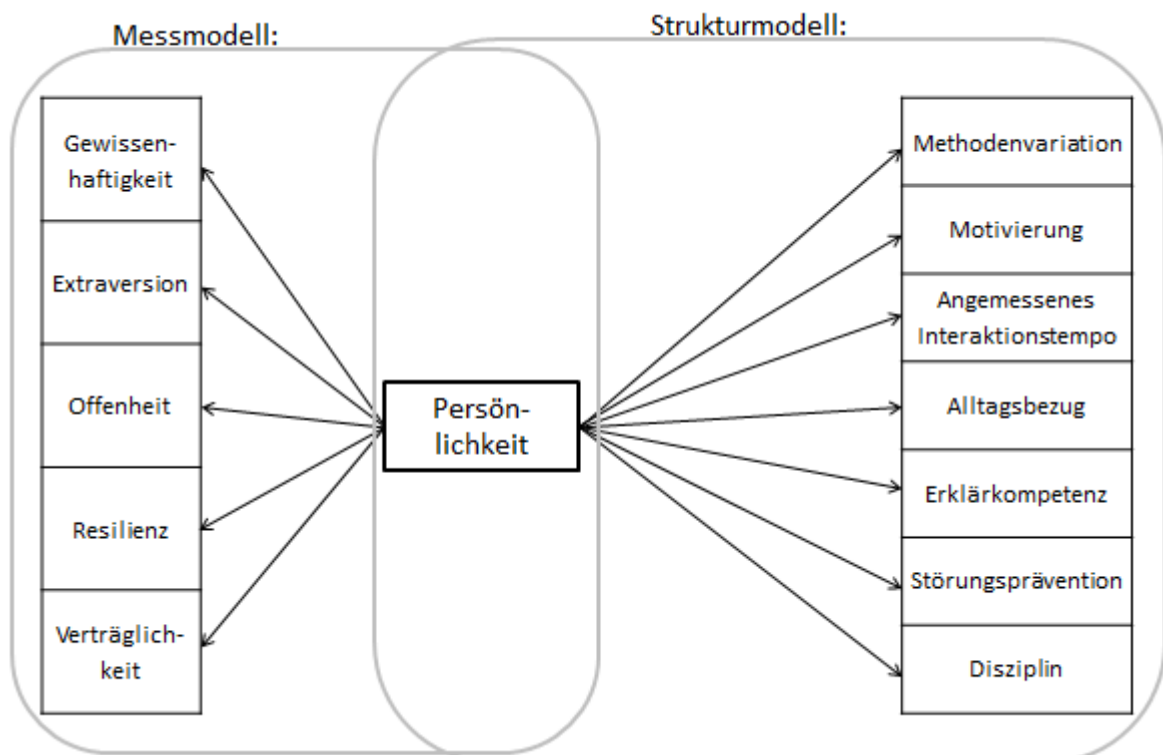


Abb. 15: Strukturgleichungsmodell mit Effekten der Lehrerpersönlichkeit (als Faktor 2. Ordnung) auf die Unterrichtsqualität (Modell 3).

Im Vergleich der Modellgüteindizes der verschiedenen Modelle sowie der Interpretation ihrer Varianzaufklärungen der abhängigen Variablen können das mögliche Verzerrungsausmaß durch Multikollinearität und die *tatsächlichen* Effektstärken der Prädiktoren eingeschätzt werden.

III. Ergebnisse

14 Deskriptive Befunde

Nachdem anhand der Stichprobenbeschreibung bereits ein erster Überblick über die Datengrundlage erfolgte, soll mit dem Bericht über ausgewählte deskriptive Befunde der Einblick in die Daten vertieft werden, um im Anschluss daran die Ergebnisse im Hinblick auf die eigentlichen Forschungsfragen präzise einordnen zu können. Die Antworten der Lehrkräfte sowie der Schülerinnen und der Schüler über die Lehrerpersönlichkeit werden im Folgenden getrennt voneinander präsentiert. Nachdem jeweils zuerst die Skalenmittelwerte dargestellt und verglichen werden, werden die Ausschöpfung der jeweiligen Skalierungen (Minimum und Maximum sowie Ausreißer), die Standardabweichung (unteres und oberes Quartil) und der Median (zur Abbildung der Schiefe der Daten) der jeweiligen Skalen anhand von Boxplots dargestellt. Alle Skalen werden außerdem detailliert hinsichtlich ihrer Geschlechtsspezifität betrachtet. Zudem werden die Interkorrelationen der verschiedenen BigFive-Dimensionen sowie der Prozessmerkmale untersucht. Hinsichtlich der Schülerperspektiven müssen – den Ergebnissen der Modellprüfungen (siehe Kap. 11.3.2 und 12.3.1) entsprechend – die beiden Fächer Mathematik und Deutsch getrennt betrachtet werden.¹³²

Bei der vorliegenden Arbeit handelt es sich in weiten Teilen um eine explorative Studie, da bislang keine systematisch erhobenen Daten bezüglich der Wahrnehmung der BigFive der Lehrerpersönlichkeit durch deren Schülerinnen und Schüler vorliegen. Somit wurde bislang auch keine solcherart erhobene Lehrerpersönlichkeit mit der schülerperzipierten Unterrichtsqualität in Verbindung gebracht. Angesichts dieses explorativen Untersuchungsdesigns erübrigt sich eine vorausgehende Aufstellung gerichteter Hypothesen, vielmehr geht es darum, anhand der Ergebnisse Hypothesen zu generieren und entlang bestehender Theorien zu erklären.

Es sei hier anzumerken, dass diverse Geschlechterunterschiede in den Mittelwerten der Schülereinschätzungen bestehen: Zum einen unterscheiden sich die Angaben innerhalb der Fächer je nach Geschlecht der Schülerinnen und Schüler,

¹³² Da sich die Mittelwerte der Schülerangaben nur marginal (bei der 3. Nachkommastelle) zwischen Individualebene und Klassenebene unterscheiden, werden diese auf die 1. Nachkommastelle gerundet berichtet. Bei Betrachtung der Interkorrelationen werden beide Werte berücksichtigt.

zum anderen je nach Geschlecht der betreffenden Lehrkraft; darüber hinaus innerhalb des Geschlechts der Lehrkräfte wiederum je nach Geschlecht der Schüler/innen sowie innerhalb des Geschlechts der Schüler/innen je nach Geschlecht der jeweiligen Lehrkraft. Auf Individual- und Klassenebene fallen zudem diese Geschlechterunterschiede je nach Persönlichkeits- oder Prozessmerkmal unterschiedlich aus. Eine Untersuchung systematischer Abweichungen kann im Rahmen dieser Arbeit nicht erfolgen, offensichtliche Befunde werden im weiteren Verlauf der Arbeit beschrieben.

14.1 Selbsteinschätzung der Lehrerpersönlichkeit

Die Modellprüfungen des vorangegangenen Methodenteils (Kap. 11.4.1) legen für die Selbsteinschätzungen der BigFive der Lehrerpersönlichkeit eine fachübergreifende Betrachtung nahe. Dieser Betrachtungsweise wird aus weiteren Gründen auch im deskriptiven Teil gefolgt: Da Lehrerinnen und Lehrer in der Regel mindestens zwei Fächer unterrichten, in unserer Untersuchung aber weder fachspezifisch noch nach Angaben zu ihren weiteren Unterrichtsfächern befragt wurden, werden die Selbsteinschätzungen der BigFive der Lehrerpersönlichkeit fachübergreifend berichtet.

Allerdings werden im weiteren Verlauf dieser Arbeit Lehrer- und Schülerangaben miteinander verknüpft. Aufgrund der eindeutig fachspezifischen Schülereinschätzungen (siehe Kap. 11.3.2) müssen die entsprechenden Modelle fachspezifisch für die Fächer Deutsch und Mathematik analysiert werden. Daher werden im Anschluss an die fachübergreifenden Befunde – trotz schwacher theoretischer Fundierung – die Mittelwertunterschiede zwischen den Deutsch- und Mathematiklehrkräften berichtet sowie die deskriptiven Befunde zusätzlich fachspezifisch beschrieben.

14.1.1 Fachübergreifende Deskription



Gewissenhaftigkeit	Extraversion	Offenheit	Resilienz	Verträglichkeit
5,9	5	5,1	4,4	5,8

Abb. 16: Fachübergreifende Mittelwerte der Lehrer-BigFive aus Lehrersicht (Selbsteinschätzung) (N=103).

In den Dimensionen *Gewissenhaftigkeit* und *Verträglichkeit* schätzen sich die Lehrerinnen und Lehrer mit Mittelwerten von 5,9 und 5,8 Punkten insgesamt am höchsten sowie mit der niedrigsten Standardabweichung von 0,9 am einheitlichsten ein. Ebenfalls hoch, jedoch etwas moderater stellen sich dagegen die Mittelwerte für *Extraversion* (5,0) und *Offenheit für neue Erfahrungen* (5,2) dar; zudem sind hier höhere Standardabweichungen zu verzeichnen: 1,3 für *Extraversion* und 1,2 für *Offenheit*. Mit einem Mittelwert von 4,4 findet sich für die *Resilienz* der schwächste Wert, ebenfalls mit einer Standardabweichung von 1,3. Insgesamt spiegeln die Mittelwerte gängige Strukturen aus Bevölkerungsumfragen (vgl. Dehne/Schupp 2007, Hülshager/Specht/Spinath 2006) und Umfragen unter Studierenden (vgl. Henoch et al. 2015) wider, die ebenfalls Mittelwerte im positiv konnotierten oberen Drittel der Skala verzeichnen.

Der folgenden Abbildung sind die fachübergreifenden Verteilungen (Minimum, Maximum, Standardabweichung, Median) der selbsteingeschätzten Lehrkräfte-BigFive zu entnehmen:

Deskriptive Befunde

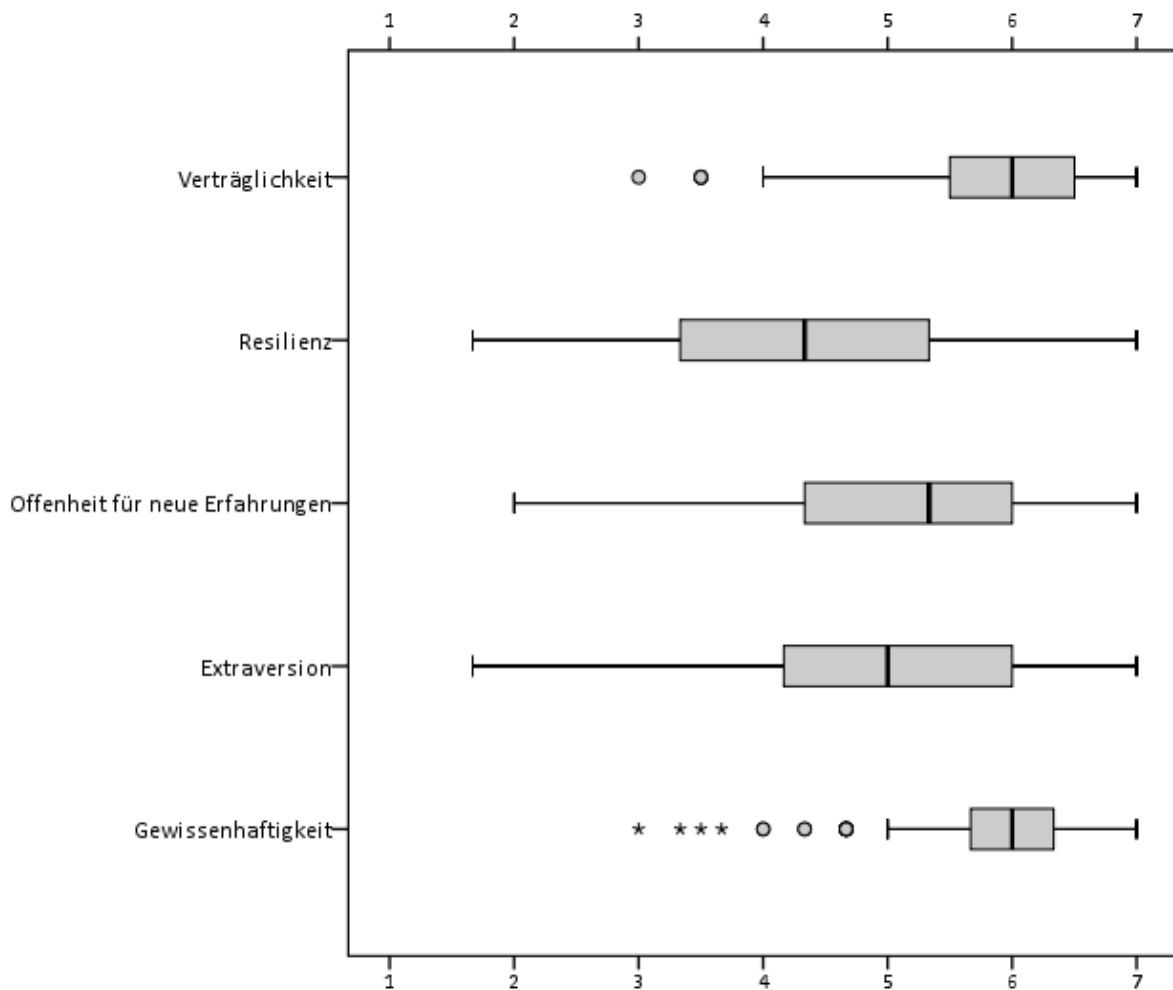


Abb. 17: Fachübergreifende Verteilung der Lehrkräfte-BigFive (Selbsteinschätzung) (N=103). Abgebildet werden die Ausschöpfung resp. Minimum und Maximum, die Standardabweichung und der Median der Skalen. Ausreißer sind als Punkte und extreme Ausreißer als Sterne abgebildet.

Es lässt sich insgesamt feststellen, dass die Skala nach oben (bis sieben) voll ausgeschöpft wurde, während niedrigere Werte unter vier in den Dimensionen *Resilienz* und *Extraversion* vorkommen. Durchweg besonders hohe Werte, mit nur vereinzelt Werten unter vier, sowie die geringste Streuung zeichnen sich in den Dimensionen *Verträglichkeit* und *Gewissenhaftigkeit* ab, bei letzterer treten Selbsteinschätzungen sogar unter fünf nur vereinzelt auf. Dementsprechend liegt neben dem oben beschriebenen Mittelwert auch der Median in diesen Dimensionen am höchsten.

Signifikante Mittelwertunterschiede ($F=15,0/p\text{-value}=0$) zwischen männlichen und weiblichen Lehrkräften finden sich lediglich in den Angaben zur Dimension *Extraversion*: Lehrerinnen schätzen sich mit einem Mittelwert von 5,6 um einen

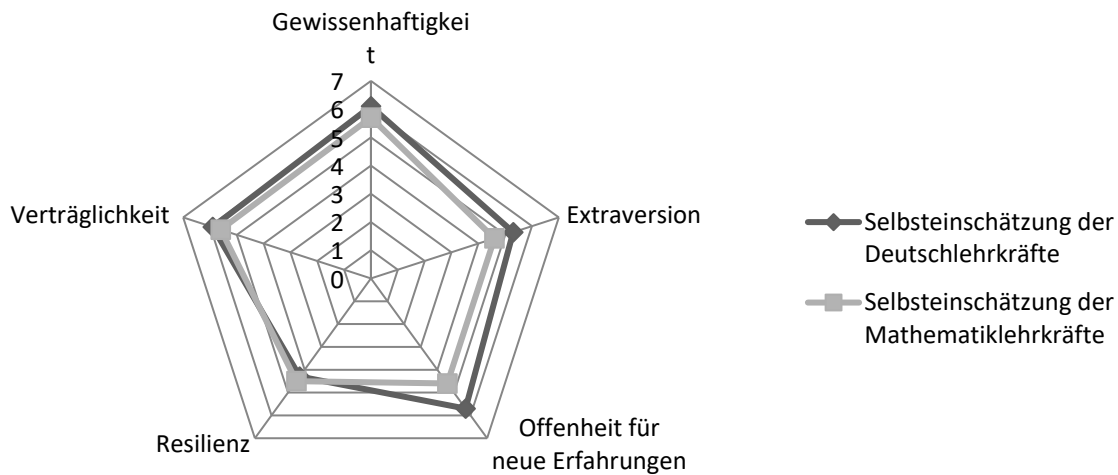
Punkt extravertierter ein als ihre Kollegen (4,6). Dies entspricht generellen geschlechtsspezifischen Befunden, denen zufolge Frauen höhere Werte in der Dimension *Extraversion* angeben. Zwar geben diesen generellen Befunden zufolge Frauen außerdem in den Dimensionen *Neurotizismus* und *Verträglichkeit* höhere Werte an als Männer, allerdings spiegelt sich diese Tendenz nicht in den vorliegenden Daten. Dass keine Unterschiede zwischen den Geschlechtern in den Dimensionen *Gewissenhaftigkeit* und *Offenheit* zu verzeichnen sind, entspricht generellen geschlechtsspezifischen Befunden (vgl. Specht/Egloff/Schmukle 2011, Dehne/Schupp 2007).

Persönlichkeitsmerkmale werden in vielen Untersuchungen hinsichtlich ihrer Prädiktionskraft für die allgemeine Lebenszufriedenheit herangezogen. Dabei lassen sich in mehreren Studien die Merkmale der Dimensionen *Extraversion* und *Resilienz* als prädiktiv wirksam ausmachen (vgl. Haller/Müller 2006). Es zeigt sich in den vorliegenden Daten, dass die *Resilienz* die Dimension mit den am schwächsten ausgeprägten Werten ist und dass die *Extraversion* je nach Geschlecht unterschiedlich ausgeprägt ist – Lehrerinnen sind demnach zufriedener mit ihrem Leben als ihre männlichen Kollegen. Hinsichtlich der Resilienz fehlen konkrete Orientierungswerte, zumal die vorliegenden Werte angesichts ihrer hohen Verzerrungsgefahr (siehe Kap. 3.4.1) mit Vorsicht zu interpretieren sind.

14.1.2 Fachspezifische Deskription

Im Anschluss an die fachübergreifende Deskription der selbsteingeschätzten Lehrkräfte-BigFive erfolgt diese fachspezifisch für die Fächer Deutsch und Mathematik.

Deskriptive Befunde



	Gewissenhaftigkeit**	Extraversion**	Offenheit**	Resilienz	Verträglichkeit
Deutsch	6,1	5,4	5,7	4,3	5,9
Mathematik	5,7	4,6	4,6	4,5	5,6

Abb. 18: Mittelwerte der BigFive-Selbsteinschätzung von Deutsch- und Mathematiklehrkräften im Vergleich ($N_{\text{Deutsch}}=53/N_{\text{Mathematik}}=50$). **Mittelwertunterschiede auf 5%-Niveau signifikant.

Signifikante (auf 5%-Niveau) Mittelwertunterschiede zwischen den Fächern bestehen hinsichtlich der Dimensionen *Gewissenhaftigkeit* ($F=10,7/p\text{-value}=0,03$), *Extraversion* ($F=5/p\text{-value}=0$) und *Offenheit für neue Erfahrungen* ($F=0,2/p\text{-value}=0$). In diesen Dimensionen schätzen sich die Deutschlehrerinnen und -lehrer deutlich höher ein als die Mathematiklehrkräfte, wobei nicht auszuschließen ist, dass v.a. in Bezug auf die Dimension *Extraversion* die Begründung nicht im Fach, sondern im ungleich höheren Anteil männlicher Mathematiklehrkräfte liegt (siehe vorherige Seite den Absatz zu Geschlechtsunterschieden). Bei Prüfung der Fach- und Geschlechtseffekte zeigt sich tatsächlich, dass die Unterschiede bzgl. *Gewissenhaftigkeit* und *Offenheit* auf das Fach, jene der *Extraversion* auf das Geschlecht zurückführen lassen. Dieser Befunde widerspricht teilweise jenem von Henoeh et al. (2015), die ebenfalls schwächere Ausprägungen ebendieser drei Dimensionen bei Studierenden technisch-mathematischer Fächer

gegenüber Studierenden sprachlich-sozialer Fächer gefunden haben, die allerdings auch bei Kontrolle des Geschlechts bestehen blieben¹³³. In der Dimension *Gewissenhaftigkeit* geben die Deutschlehrkräfte im Mittel 6,1 Punkte an, gegenüber 5,7 Punkten der Mathematiklehrkräfte; zumal die Standardabweichung innerhalb der Gruppe der Deutschlehrerinnen und -lehrer lediglich bei 0,7 und bei den Mathematiklehrerinnen und -lehrern bei 1 liegt. Mit 5,4 im Mittel (SD=1) schätzen sich die Lehrkräfte in Deutsch, die Lehrkräfte im Fach Mathematik hingegen mit 4,6 (SD=1,4) in der Dimension *Extraversion* ein. Als *offen für neue Erfahrungen* schätzen sich die Deutschlehrkräfte mit 5,7 (SD=1) ein, während der Mittelwert der Mathematiklehrkräfte bei 4,6 (SD=1,1) liegt.

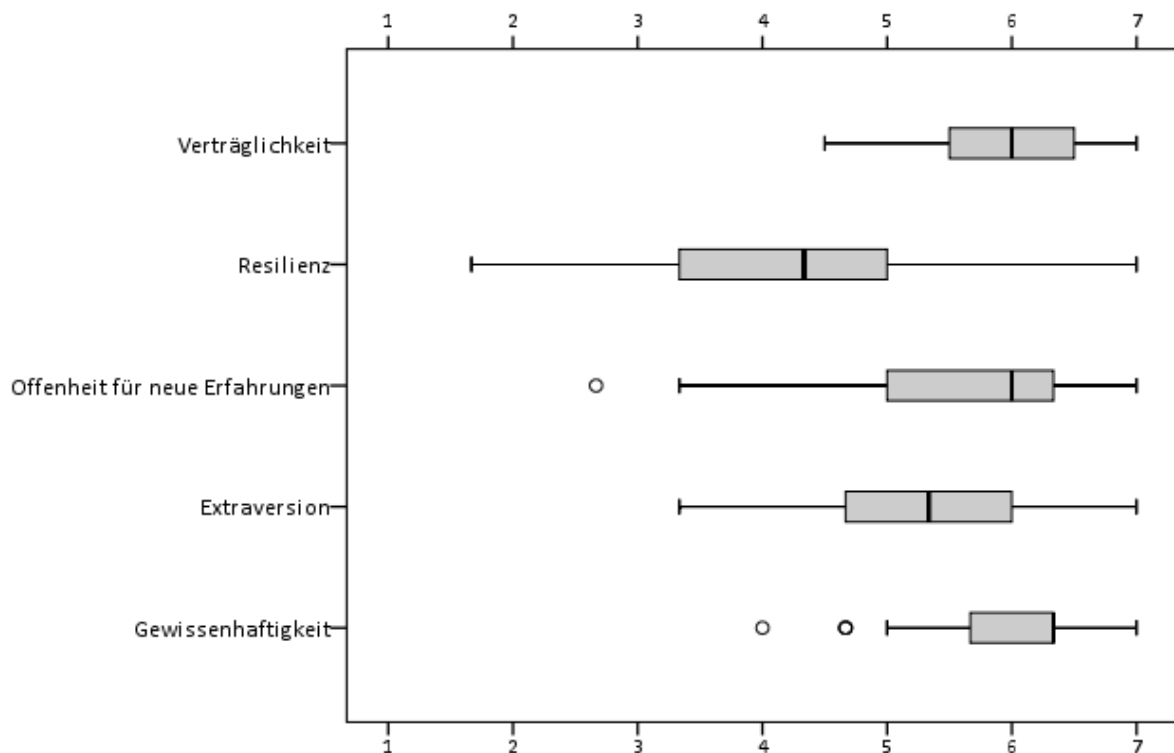


Abb. 19: Verteilungen der Lehrkräfte-BigFive (Selbsteinschätzung) im Fach Deutsch (N=53).

Abgebildet werden die Ausschöpfung resp. Minimum und Maximum, die Standardabweichung und der Median der Skalen. Ausreißer sind als Punkte abgebildet.

¹³³ Darüber hinaus zeigt die Studie, dass unter Studierenden sogenannter STEM-Fächer (science, technology, engineering, and mathematics) die Lehramtsanwärter insgesamt signifikant extravertierter sind als ihre Kommilitonen. Außerdem verzeichnet die Studie nicht-signifikante Unterschiede in der Dimension Resilienz resp. Neurotizismus, die ebenfalls den in der vorliegenden Untersuchung gefundenen fachspezifischen Tendenzen entsprechen (vgl. Henoch et al. 2015).

Deskriptive Befunde

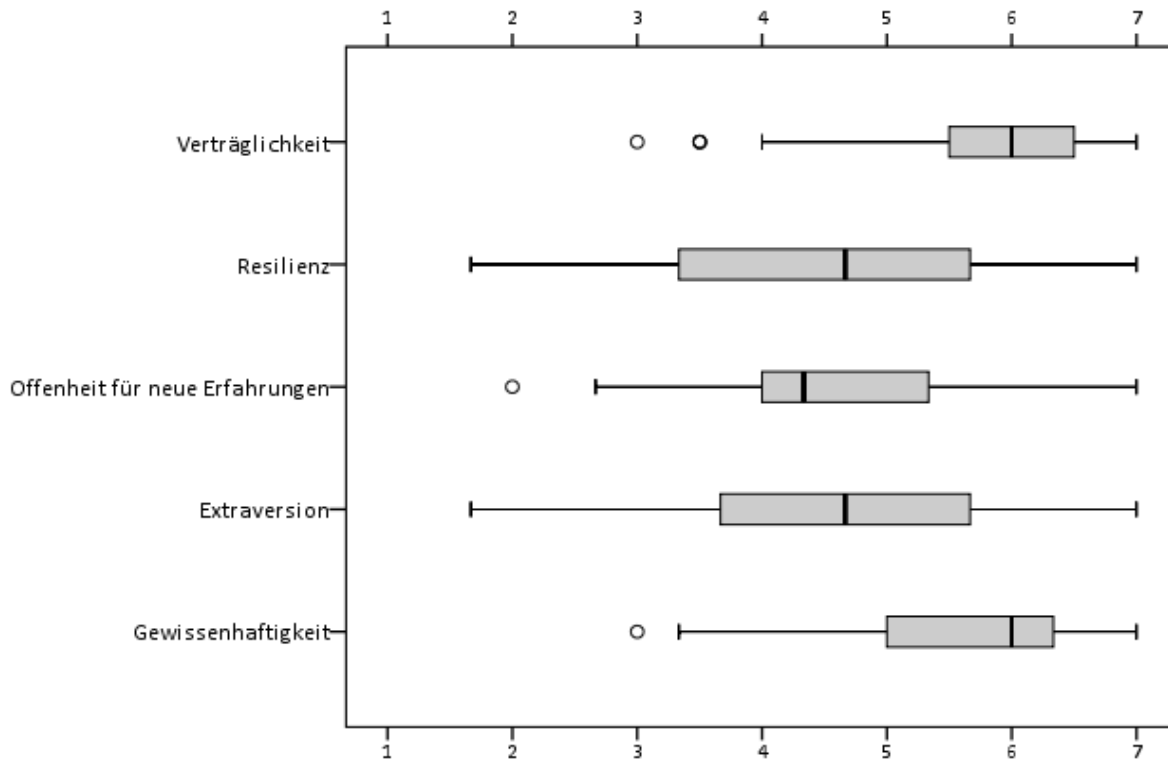


Abb. 20: Verteilungen der Lehrkräfte-BigFive (Selbsteinschätzung) im Fach Mathematik ($N=50$). Abgebildet werden die Ausschöpfung resp. Minimum und Maximum, die Standardabweichung und der Median der Skalen. Ausreißer sind als Punkte abgebildet.

14.1.3 Interkorrelationen zwischen den von den Lehrkräften selbsteingeschätzten BigFive-Dimensionen

Die Dimension *Extraversion* korreliert mit den beiden Dimensionen *Offenheit für neue Erfahrungen* und *Resilienz* am deutlichsten und auf 5%-Niveau signifikant, mit der Dimension *Verträglichkeit* hingegen schwach (und nicht signifikant). Dagegen korreliert *Resilienz* kaum mit *Offenheit* und noch weniger mit *Verträglichkeit*. Alle anderen Dimensionen korrelieren moderat auf 5%-Niveau miteinander (siehe folgende Tabelle).

Tab. 21: Interkorrelationen der BigFive-Dimensionen aus Lehrersicht (Selbsteinschätzung) mit Signifikanz auf 5%-Niveau (N=103).

	Gewissenhaftigkeit	Extraversion	Offenheit	Resilienz
Extraversion	,20*			
Offenheit	,23**	,41**		
Resilienz	,22**	,32**	,12	
Verträglichkeit	,20*	,17	,24**	-,05

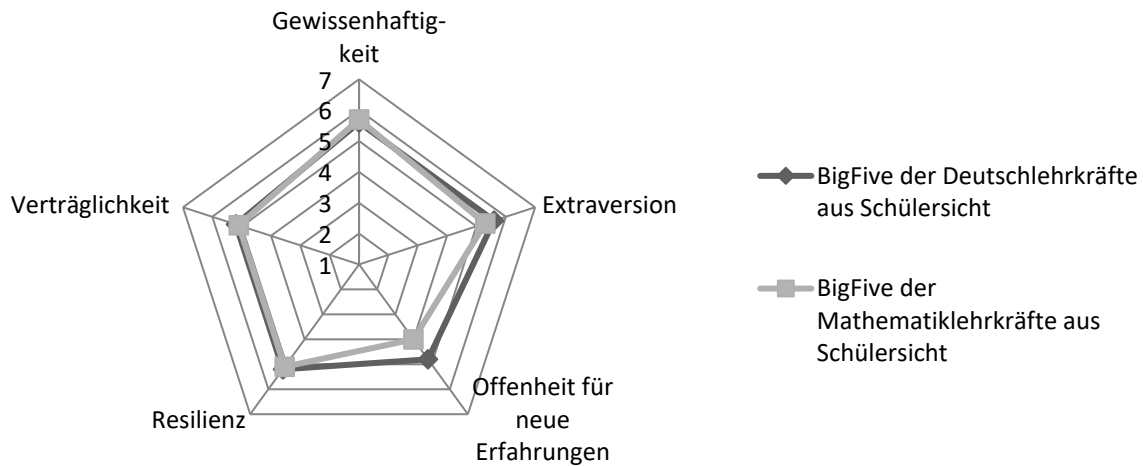
Die deutlichen Interkorrelationen zwischen Extraversion und Offenheit ($r=0,41$) bestätigen den Befund u.a. von Lang, Lüdtke und Asendorpf (2001), die zudem auch von insgesamt recht ausgeprägten Faktorkorrelationen der BigFive berichten. Hülshager et al. (2006) verzeichnen in ihrer Untersuchung des NEO-PI-R außerdem erhöhte negative Korrelationen zwischen Neurotizismus und Extraversion sowie Gewissenhaftigkeit, was das positive Korrelationsniveau der vorliegenden Untersuchung zwischen Resilienz und den anderen Dimensionen spiegelt. Die Interkorrelationen zwischen den BigFive-Dimensionen widersprechen zwar einer statistisch unabhängigen Erhebung der BigFive-Dimensionen. Womöglich sind sie auf dieselbe adjektivistische Erhebungsart zurückzuführen, die generell anfällig für Bias durch individuelle Antworttendenzen ist (vgl. Lang/Lüdtke/Asendorpf 2001).

14.2 Schülereinschätzung der Lehrerpersönlichkeit

Wie die Modellprüfungen in Kap. 11.3.2 nahelegen, schätzen die Schülerinnen und Schüler ihre Lehrkräfte fachspezifisch ein. Dementsprechend werden diese Befunde ebenfalls fachspezifisch dargestellt und verglichen. Darüber hinaus zeigen sich signifikante Mittelwertunterschiede zwischen den Fächern hinsichtlich der Dimensionen *Extraversion* und *Offenheit für neue Erfahrungen*, in der Dimension *Resilienz* wird die Signifikanz auf 1%-Niveau knapp verfehlt ($p\text{-value}=0,012$). In den Dimensionen *Gewissenhaftigkeit* und *Verträglichkeit* lassen sich keine signifikanten Fachunterschiede der Mittelwerte feststellen (4,6% und 2,1%).

Auffällig ist, dass von den Schülerinnen und Schülern in allen Dimensionen und beiden Fächern die Skalierung (von 1-7) voll ausgeschöpft wurde.

Deskriptive Befunde



	Gewissenhaftigkeit	Extraversion**	Offenheit**	Resilienz ⁺	Verträglichkeit
Deutsch	5.6	5.6	4.8	5.2	5.2
Mathematik	5.7	5.3	4.0	5.1	5.1

Abb. 21: Mittelwerte der BigFive der Deutsch- ($N=2235-2237$) und Mathematiklehrkräfte ($N=2230-2239$) aus Schülersicht (Fremdeinschätzung). **Mittelwertunterschiede auf 1%-Niveau signifikant.

Die (signifikanten) Unterschiede der Schülereinschätzungen zwischen den Deutsch- und Mathematiklehrkräften in den Dimensionen *Extraversion* und *Offenheit für Erfahrungen* ließen sich einerseits durch ein tatsächliches, möglicherweise fachlich bedingtes unterschiedliches Auftreten der jeweiligen Fachlehrerinnen und -lehrer erklären und andererseits durch eine höhere Kompatibilität der Itemformulierungen mit einem der beiden Fächer. Hinsichtlich Letzterem ließen sich die (zwei der insgesamt drei) Attributionen „lebhaftes Phantasie, Vorstellung“ und „künstlerische ästhetische Erfahrungen“ in der Dimension *Offenheit* plausibel eher auf Inhalte des Deutschunterrichts beziehen als auf jene des Mathematikunterrichts. Entsprechend ist in der Dimension *Extraversion* v.a. die Itemformulierung mit der Attribution „kommunikativ, gesprächig“ auf Kommunikation ausgerichtet, was eher dem Fach Deutsch entspräche.

14.2.1 Deutsch

Im Fach Deutsch schätzen die Schülerinnen und Schüler ihre Lehrerinnen und Lehrer in den Dimensionen *Gewissenhaftigkeit* und *Extraversion* mit einem Mittelwert von 5,6 am höchsten ein; zudem liegt in diesen beiden Dimensionen die Standardabweichung mit 1,3 und 1,2 am niedrigsten und deutet damit auf ein relativ einheitliches Antwortverhalten hin. Am zweithöchsten mit Mittelwerten von jeweils 5,2 Punkten schätzen die Schülerinnen und Schüler ihre Deutschlehrkräfte in den Dimensionen *Resilienz* und *Verträglichkeit* ein, beides mit Standardabweichungen von 1,4. Der niedrigste Mittelwert liegt in der Dimension *Offenheit für neue Erfahrungen* mit 4,8 Punkten und ebenfalls einer Standardabweichung von 1,4.

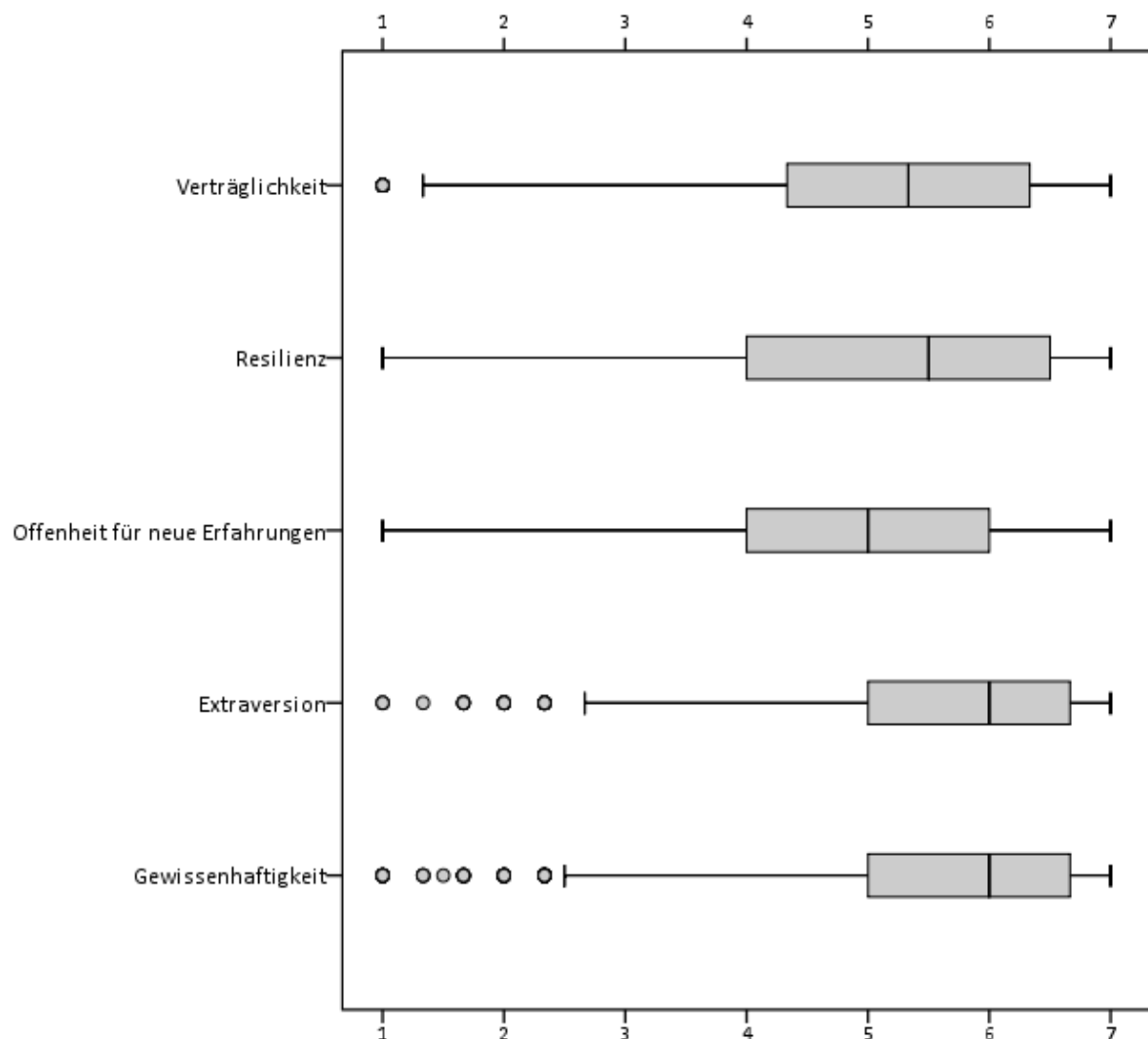


Abb. 22: Häufigkeitsverteilung der Deutschlehrkräfte-BigFive aus Schülersicht (Fremdeinschätzung; N=2230-2239). Abgebildet werden die Ausschöpfung resp. Minimum und Maximum, die Standardabweichung und der Median der Skalen. Ausreißer sind als Punkte abgebildet.

14.2.2 Mathematik

Im Fach Mathematik führt die Dimension *Gewissenhaftigkeit* mit einem Mittelwert von 5,7 die Schülereinschätzungen der BigFive ihrer Lehrkräfte an. Die Standardabweichung entspricht mit 1,3 der der Dimension *Extraversion*, die einen Mittelwert von 5,3 ausweist. Die Dimensionen *Verträglichkeit* und *Resilienz* zeigen sich vergleichbar wie im Fach Deutsch mit Mittelwerten von 5,1 Punkten und Standardabweichungen von 1,5. Die Dimension *Offenheit für neue Erfahrungen* bei ihren Mathematiklehrerinnen und -lehrern wird von den Schülerinnen und Schülern mit einem Mittelwert von 4,0 eingeschätzt (und einer Standardabweichung von 1,5).

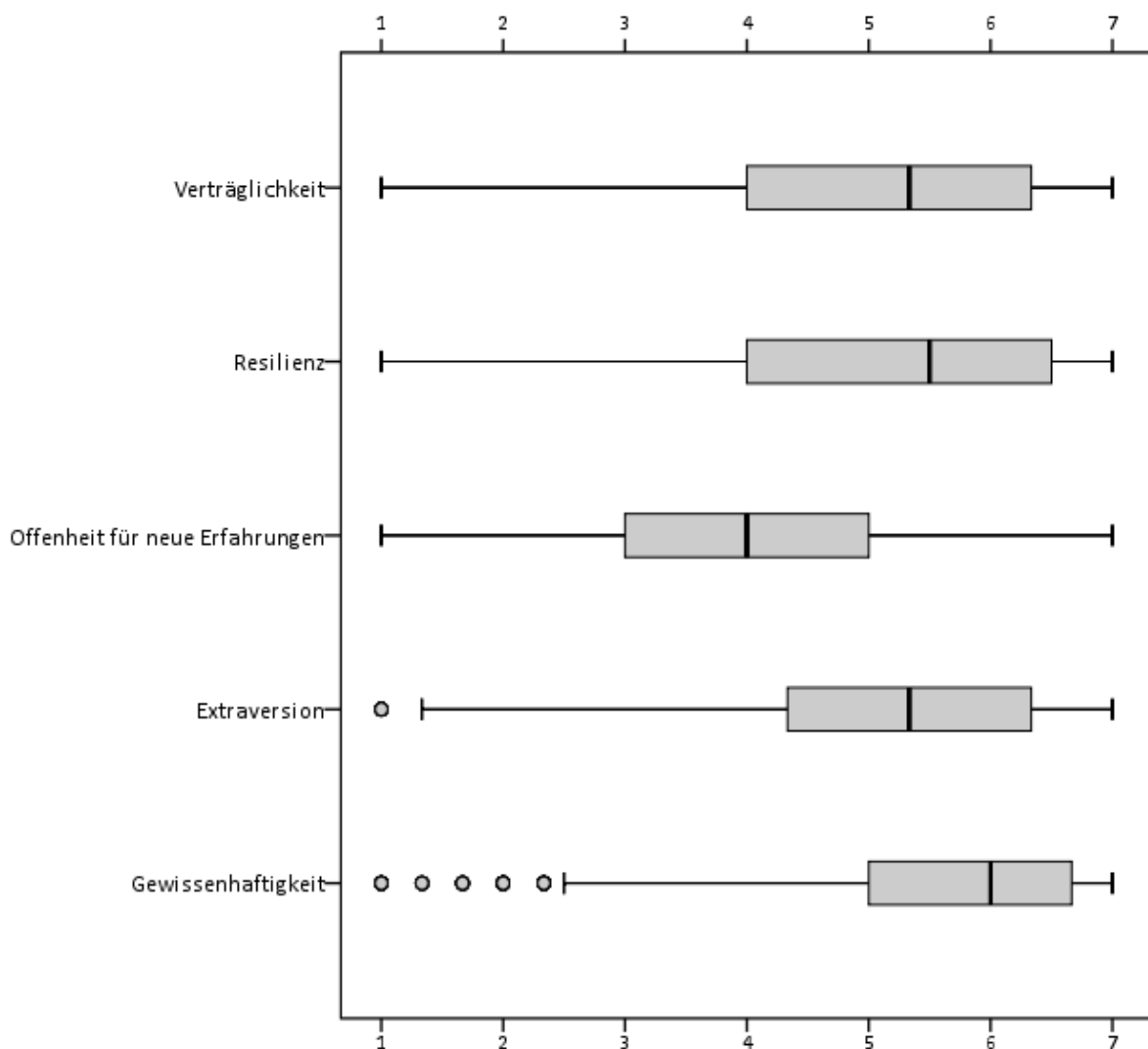


Abb. 23: Häufigkeitsverteilung der Lehrkräfte-BigFive aus Schülersicht (Fremdeinschätzung; $N=2230-2239$). Abgebildet werden die Ausschöpfung resp. Minimum und Maximum, die Standardabweichung und der Median der Skalen. Ausreißer sind als Punkte abgebildet.

14.2.3 Geschlechterdifferenzen innerhalb der Fächer

Im Fach Deutsch unterscheiden sich Schülerinnen und Schüler in ihren Angaben signifikant ($F=35,7/p\text{-value}=0$) geschlechtsspezifisch lediglich in der Dimension *Offenheit für neue Erfahrungen*: Hier liegen die Schüler bei einem Mittelwert von 4,9 und die Schülerinnen bei einem Mittelwert von 4,6.

Signifikante Geschlechterdifferenzen zeigen sich im Fach Mathematik hingegen in den Angaben zu den Dimensionen *Offenheit für neue Erfahrungen* ($F=22,2/p\text{-value}=0$) und *Verträglichkeit* ($F=11,1/p\text{-value}=0$). Während die Schülerinnen ihre Lehrkräfte in der Dimension *Offenheit für neue Erfahrungen* mit einem Mittelwert von 3,8 wahrnehmen, schreiben Schüler ihnen im Mittel einen Wert von 4,1 zu. Hinsichtlich der Dimension *Verträglichkeit* dagegen schreiben die Schülerinnen ihren Lehrkräften mit 5,2 den höheren Mittelwert zu als die Schüler mit 5,0.

Wie zu Anfang des Ergebnisteils erwähnt, können nicht alle Geschlechterdifferenzen in den Skalenmittelwerten analysiert werden, da sich die Abweichungen als komplexes Muster darstellen. Dennoch wird vollständigkeithalber¹³⁴ berichtet, inwiefern die Mittelwerte der Schülerinnen- und Schülerangaben innerhalb der jeweiligen Fächer je nach Geschlecht der Lehrkraft variieren: In Mathematik zeigen sich signifikante (5%-Niveau) Mittelwertunterschiede der Schülereinschätzungen je nach Geschlecht der Lehrkraft innerhalb der Dimension *Resilienz*. Hier werden die Lehrer mit 5,3 Punkten resilienter eingeschätzt als die Lehrerinnen mit 4,7 Punkten. In der Dimension *Offenheit* wird eine Signifikanz der Mittelwertunterschiede knapp verfehlt – hier werden Lehrer mit 4,1 und Lehrerinnen mit 3,6 eingeschätzt. In den anderen Dimensionen sowie im Fach Deutsch lassen sich keine berichtenswerten Mittelwertunterschiede feststellen.

14.2.4 Interkorrelationen zwischen den schülerperzipierten BigFive-Dimensionen der Lehrkräfte

Alle Dimensionen korrelieren im Fach Deutsch und im Fach Mathematik sowie auf Individual- als auch auf Klassenebene signifikant und hoch miteinander. Aufgrund der Vielzahl an Parametern auf latenter Ebene ist die Komplexität eines

¹³⁴ Im weiteren Verlauf der Arbeit werden Geschlechtseffekte nur dann kontrolliert, wenn die Schüler- und Lehrerdaten zusammengefasst werden (Forschungsfrage 1), da sich ansonsten die Stichprobengröße deutlich unterscheidet (siehe Kap. 6.3).

Mehrebenenmodells zu hoch, daher werden innerhalb der Interkorrelationsmatrix die manifesten Skalenmittelwerte eingesetzt.¹³⁵

Tab. 22: *Interkorrelationen der BigFive-Dimensionen aus Schülersicht (alle Dimensionen interkorrelieren signifikant auf beiden Ebenen; Signifikanz_{within}=0,01/Signifikanz_{between}=0,05; N_{within}=2230-2239/N_{between}=123).*

	Gewissenhaftigkeit	Extraversion	Offenheit	Resilienz
Deutsch (within/between)				
Extraversion	,45**/,56**			
Offenheit	,42**/,62**	,46**/,83**		
Resilienz	,39**/,57**	,37**/,67**	,27**/,63**	
Verträglichkeit	,43**/,47**	,35**/,37**	,32**/,65**	,42**/,51**
	Gewissenhaftigkeit	Extraversion	Offenheit	Resilienz
Mathematik (within/between)				
Extraversion	,41**/,58**			
Offenheit	,29**/,63**	,44**/,86**		
Resilienz	,38**/,60**	,37**/,74**	,24**/,68**	
Verträglichkeit	,43**/,57**	,35**/,54**	,32**/,69**	,40**/,51**

Es wird deutlich, dass in beiden Fächern v.a. die Dimensionen *Extraversion* und *Offenheit* auf Klassenebene sehr hoch korreliert sind. Als Maß für eine mögliche Kollinearität wird in der Literatur $r > 0,9$ angesetzt (vgl. Stapleton et

¹³⁵ Bei stichpunktartigen Überprüfungen zeigt sich, dass die Korrelationen auf latenter Ebene tendenziell höher ausfallen als auf manifester Ebene.

al. 2016, Allison 2012), das hier nur knapp verfehlt wird. Daher ist in den Regressionsanalysen zu Forschungsfrage 2 mit durch Multikollinearität verzerrten Regressionskoeffizienten zu rechnen. Dieser Befund wird in den multiplen Regressionsmodellen zu Forschungsfrage 2 berücksichtigt, indem die Korrelation zwischen *Extraversion* und *Offenheit* mit in die Modelle aufgenommen wird.

14.3 Unterrichtsqualität aus Schülersicht

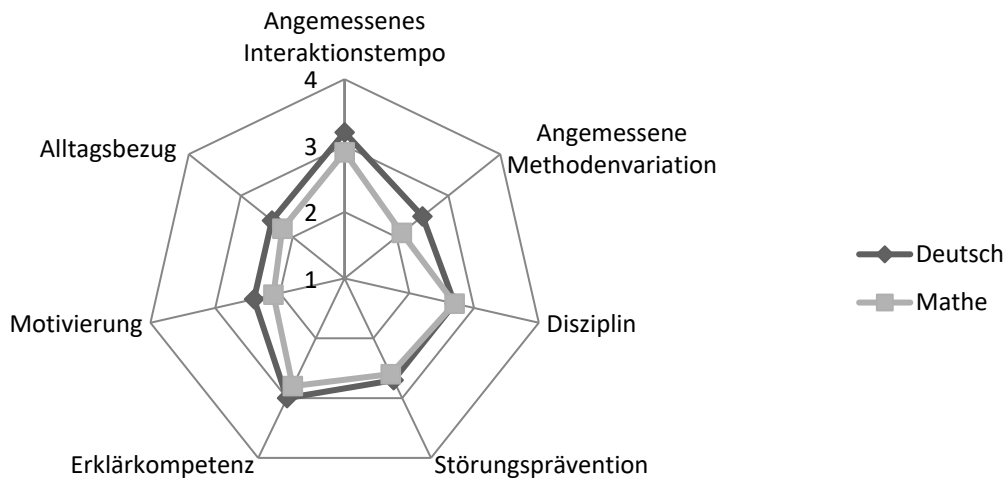
Die Güte der Modellprüfungen zur schülerperzipierten Unterrichtsqualität anhand von Prozessmerkmalen des Unterrichts zeigte deutlich, dass zwischen verschiedenen Fächern differenziert werden muss (vgl. Kap. 12.3.1). Daher wird die schülerperzipierte Unterrichtsqualität der beiden Fächer Mathematik und Deutsch getrennt berichtet. Zudem weisen – außer das Prozessmerkmal *Disziplin* – alle Prozessmerkmale der Unterrichtsqualität signifikante Mittelwertunterschiede zwischen den Fächern auf¹³⁶.

Darüber hinaus ist auffällig, dass das Fach Mathematik gegenüber dem Fach Deutsch insgesamt auf niedrigerem Skalenniveau eingeschätzt wird. Obgleich in beiden Fächern die gesamte Skalenbreite (4-stufige Likertskalierung) von den Schülerinnen und Schülern ausgeschöpft wurde. Dies entspricht generellen Forschungsbefunden aus der Unterrichtsforschung, denen zufolge Themengebiete wie Mathematik und Naturwissenschaften stets tendenziell niedrige Ratingwerte verzeichnen (vgl. Ramsden 1991), woraus sich ergibt, dass daher keine fachübergreifenden Vergleiche möglich sind.

Insgesamt lassen die Schülereinschätzungen auf eine positive Wahrnehmung des Unterrichts schließen, was generellen Forschungsbefunden entspricht (vgl. Ditton/Arnold 2004, Seifried 2006), allerdings spiegelt die schwach ausgeprägte Methodenvariation vor allem in Mathematik den weitgehend gesicherten Forschungsbefund hinsichtlich der Variationsarmut eingesetzter Unterrichtsmethoden wider (vgl. Ditton 2002).

¹³⁶ Dies wurde auf Klassenebene ermittelt; auf Individualebene zeigen sich zudem moderat signifikante Mittelwertunterschiede hinsichtlich des Merkmals *Störungsprävention*.

Deskriptive Befunde



	Interaktions-tempo**	Metho-denvaria-tion**	Disziplin	Störungs-präven-tion*	Erklär-kompe-tenz**	Motivie-rung**	Alltags-bezug**
Deutsch	3,2	2,5	2,7	2,7	3	2,4	2,4
Mathe-matik	2,9	2,1	2,7	2,6	2,8	2,1	2,2

Abb. 24: Mittelwerte der Prozessmerkmale der Unterrichtsqualität aus Schülersicht im Fach Deutsch (N=2230-2240) und im Fach Mathematik (N=2229-2244). ***/**Mittelwertunterschiede auf 1%-Niveau signifikant.

14.3.1 Deutsch

Es lässt sich feststellen, dass der Deutschunterricht insgesamt eher positiv wahrgenommen zu werden scheint: die Skalenwerte befinden sich bei allen Prozessmerkmalen im Bereich der positiv konnotierten Beurteilungswerte. Zudem lassen sich keine Mittelwertunterschiede zwischen Individual- und Klassenebene feststellen.

Der niedrigste Wert von 2,4 Punkten findet sich in den beiden Merkmalen *Alltagsbezug* und *Motivierungsqualität*, bei beiden mit einer Standardabweichung von 0,8 – hier bestehen zudem signifikante Geschlechterunterschiede: Während Jungen die *Motivierungsqualität* etwas positiver einschätzen als Mädchen ($F=8,6/p\text{-value}=0$), dreht sich dieses Verhältnis hinsichtlich des *Alltagsbezugs* genau um ($F=9,2/p\text{-value}=0$). Im Mittelfeld der Beurteilung liegen die Angemessenheit der

Deskriptive Befunde

Methodenvariation mit 2,5 (SD=0,8) sowie die *Störungsprävention* und die *Disziplin*, beide mit 2,7 Punkten und Standardabweichungen von 0,7. In Bezug auf die Einschätzung der Angemessenheit der *Methodenvariation* lässt sich ebenfalls ein signifikanter Unterschied zwischen den Geschlechtern feststellen, wobei dieser mit 0,1 Punkten Unterschied (Schüler liegen in ihrer Einschätzung leicht über derjenigen der Schülerinnen) eher schwach ausfällt. Am positivsten werden von den Schülerinnen und Schülern die Prozessmerkmale *Erklärkompetenz* mit 3,0 (SD=0,7) und *Angemessenes Interaktionstempo* mit 3,2 Punkten (SD=0,6) beurteilt. Es lässt sich die Tendenz beobachten, dass mit zunehmend positiver Einschätzung die Standardabweichung abfällt.

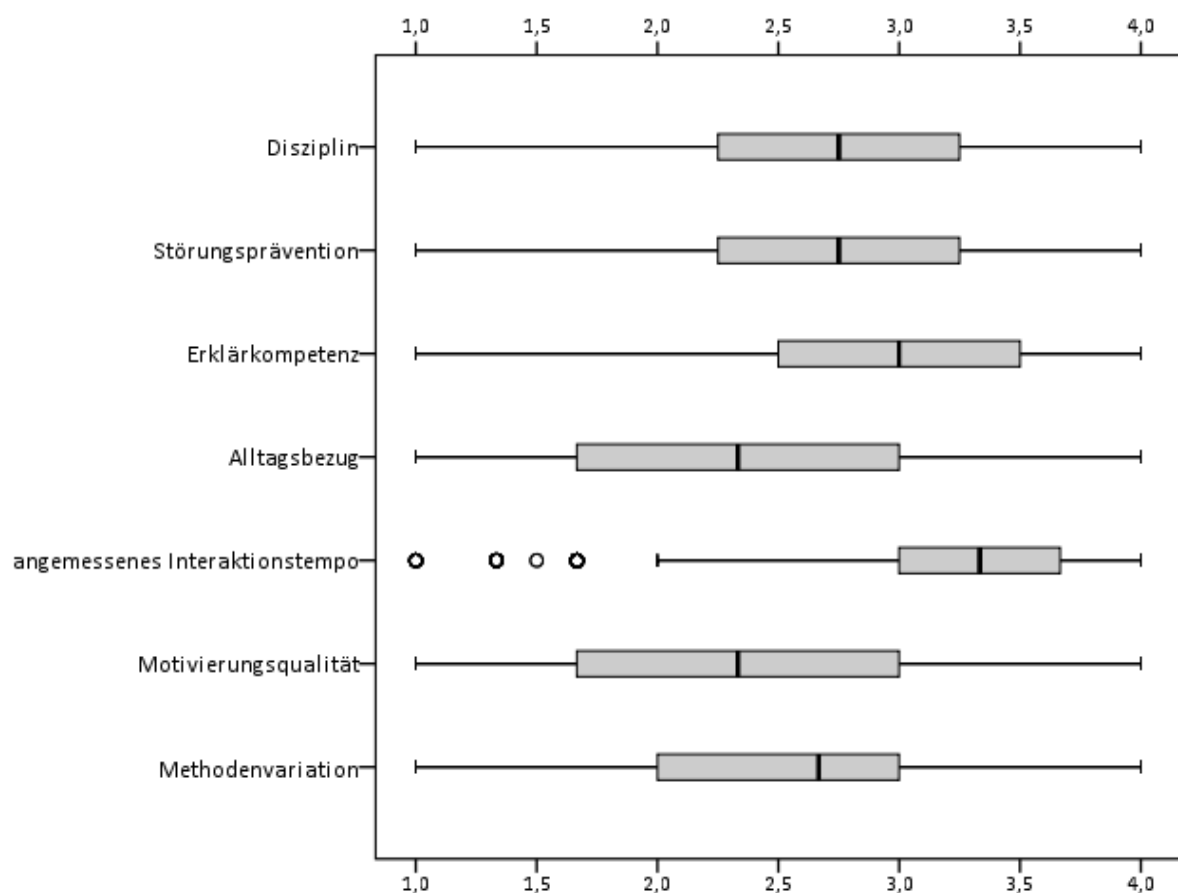


Abb. 25: Häufigkeitsverteilung der Prozessmerkmale des Deutschunterrichts aus Schülersicht (N=2230-2240). Abgebildet werden die Ausschöpfung resp. Minimum und Maximum, die Standardabweichung und der Median der Skalen. Ausreißer sind als Punkte abgebildet.

14.3.2 Mathematik

Die Schülereinschätzungen der Unterrichtsqualität im Fach Mathematik rangieren zwischen 2,1 und 2,9 Skalenpunkten, liegen also insgesamt tendenziell im guten Mittelfeld. Es lassen sich keine Mittelwertunterschiede zwischen Individual- und Klassenebene feststellen.

Am schwächsten werden die *Angemessenheit der Methodenvariation*¹³⁷ und die *Motivierungsqualität* mit jeweils 2,1 Punkten wahrgenommen, direkt gefolgt vom Merkmal *Alltagsbezug* (2,2). Auch hier zeigen sich signifikante Geschlechterunterschiede (alle mit p-value=0): die Schülerinnen vergeben bei der Einschätzung dieser Merkmale durchschnittlich 0,2 Punkte weniger als ihre Mitschüler.¹³⁷ Die Merkmale *Störungsprävention* (2,6) und *Disziplin* (2,7) liegen im Fach Mathematik ebenso wie im Fach Deutsch im Mittelfeld der Schülerbeurteilung. Doch hier findet sich hinsichtlich der Einschätzung der *Disziplin* ein kleiner, jedoch signifikanter Unterschied zwischen den Schülerinnen und Schülern: Schülerinnen schätzen diese mit 2,7 und Schüler mit 2,8 ein (p-value=0,01). Wie im Fach Deutsch weisen die Merkmale *Klarheit und Strukturiertheit* mit 2,8 und vor allem das *Angemessene Interaktionstempo* mit 2,9 die höchsten Werte auf.

¹³⁷ Alltagsbezug: F=22,5; Motivierungsqualität: F=20,1; Methodenvariation: F=15,2

Deskriptive Befunde

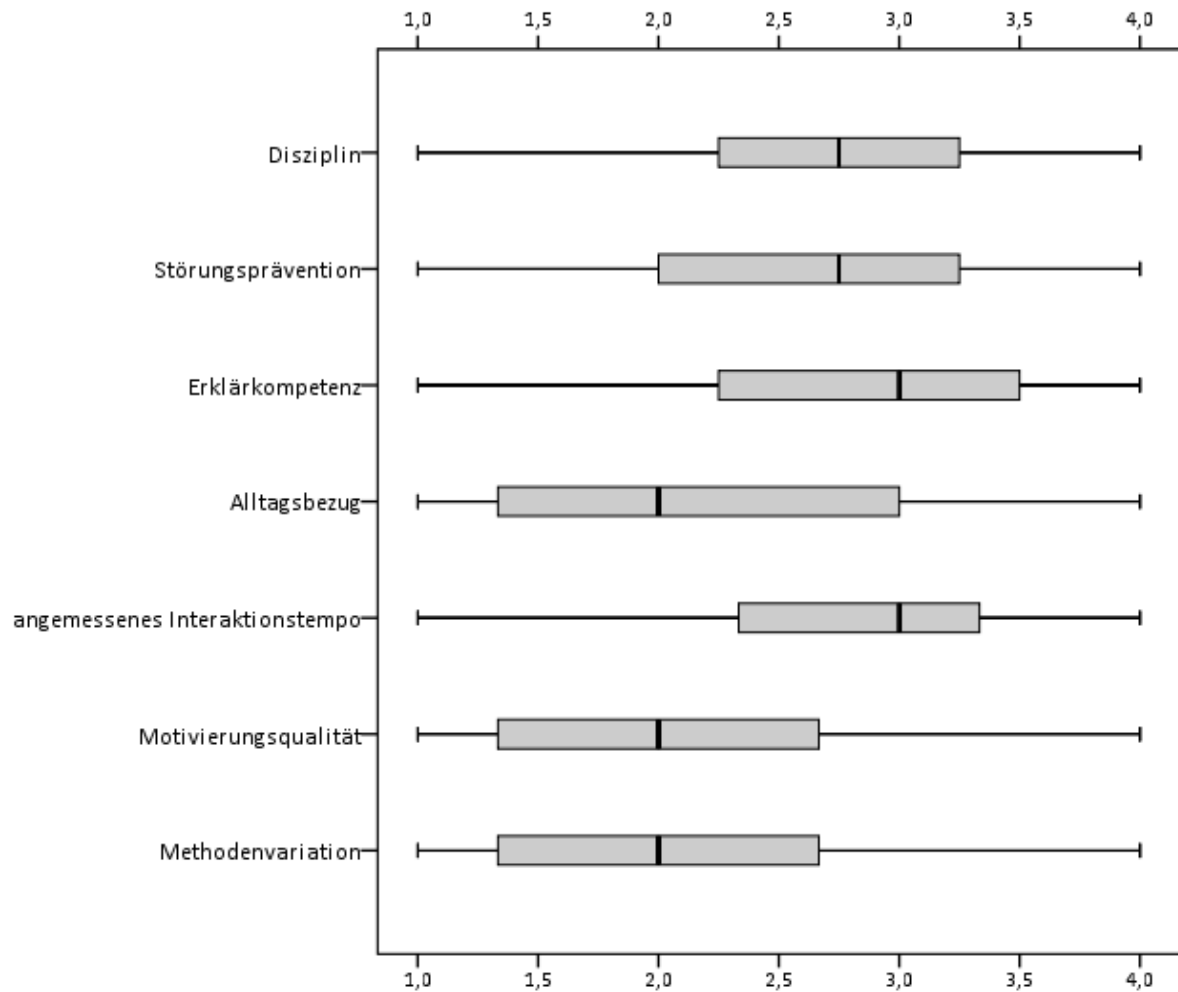


Abb. 26: Häufigkeitsverteilung der Prozessmerkmale des Mathematikunterrichts aus Schülersicht ($N=2229-2244$). Abgebildet werden die Ausschöpfung resp. Minimum und Maximum, die Standardabweichung und der Median der Skalen.

14.3.3 Interkorrelationen zwischen den schülerperzipierten Prozessmerkmalen der Unterrichtsqualität

Alle Prozessmerkmale korrelieren innerhalb der beiden Fächer Deutsch und Mathematik signifikant miteinander (siehe folgende Tabelle):

Deskriptive Befunde

Tab. 23: Interkorrelationen zwischen den Prozessmerkmalen der Unterrichtsqualität nach Fächern (**Signifikanz_{within}=0,01/Signifikanz_{between}=0,05; $N_{within}=2229-2244/N_{between}=123$).

	Metho- den- variation	Motivie- rungs- qualität	Interakti- onstempo	Alltags- bezug	Erklär- kompe- tenz	Störungs- präven- tion
Deutsch (within/between)						
Motivie- rungs- qualität	,45**/ ,63**					
Interakti- onstempo	,26**/ ,42**	,30**/ ,54**				
Alltags- bezug	,27**/ ,30**	,59**/ ,72**	,18**/ ,28**			
Erklär- kompe- tenz	,41**/ ,58**	,63**/ ,87**	,39**/ ,60**	,42**/ ,61**		
Störungs- präven- tion	,29**/ ,38**	,41**/ ,54**	,11**/ ,10	,32**/ ,55**	,48**/ ,62**	
Disziplin	,28**/ ,43**	,35**/ ,57**	,18**/ ,22**	,25**/ ,49**	,40**/ ,59**	,43**/ ,61**
Mathematik (within/between)						
	Metho- den- variation	Motivie- rungs- qualität	Interakti- onstempo	Alltags- bezug	Erklär- kompe- tenz	Störungs- präven- tion
Motivie- rungs- qualität	,43**/ ,71**					
Interakti- onstempo	,26**/ ,57**	,37**/ ,63**				
Alltags- bezug	,33**/ ,62**	,58**/ ,71**	,26**/ ,50**			
Erklär- kompe- tenz	,39**/ ,65**	,66**/ ,88**	,49**/ ,70**	,43**/ ,63**		
Störungs- präven- tion	,22**/ ,29**	,35**/ ,44**	,09**/ ,05	,23**/ ,27**	,41**/ ,47**	
Disziplin	,24**/ ,43**	,42**/ ,66**	,24**/ ,30**	,25**/ ,45**	,44**/ ,64**	,37**/ ,48**

Deskriptive Befunde

Insgesamt lässt sich feststellen, dass sich die Tendenzen auf Individualebene entsprechend auf der Klassenebene abzeichnen. Am stärksten korrelieren in beiden Fächern die Merkmale *Motivierungsqualität* und *Klarheit/Strukturiertheit*. Die schwächsten und einzigen (auf Klassenebene) nicht signifikanten Korrelationen lassen sich sowohl in Deutsch als auch in Mathematik zwischen dem *Angemessenen Interaktionstempo* und dem Merkmal *Störungsprävention* ausmachen.

15 Forschungsfrage 1: Welcher Zusammenhang besteht zwischen selbst- und fremd-wahrgenommener Lehrerpersönlichkeit?

Der erste Analyseschritt in dieser Arbeit besteht darin, die Idee einer mit sich identischen Lehrerpersönlichkeit aus Innen- wie aus Außenbetrachtung zu prüfen. Dazu werden die Fremdwahrnehmungen der Lehrerpersönlichkeit durch die Schülerinnen und Schüler den Selbstwahrnehmungen ihrer Lehrerinnen und Lehrer gegenübergestellt und entsprechende Korrelationen analysiert. Dies erfolgt – gemäß den Befunden aus den Messmodellen der Schülerantworten (Kap. 11.3.2) – fachspezifisch in den Fächern Deutsch und Mathematik. Bereits hier lassen sich die ersten Differenzen in den Wahrnehmungen je nach Perspektive ausmachen: während einerseits die Lehrerinnen und Lehrer keine fachspezifischen Selbsteinschätzungen ablegen (da sie mehrere Fächer unterrichten und eher ausgeschlossen ist, dass sie ihre Persönlichkeit je nach Fach ändern), wurden andererseits die Schülerinnen und Schüler explizit nach ihrer Wahrnehmung der jeweiligen Fachlehrerinnen und Fachlehrer in Deutsch und Mathematik befragt. Dementsprechend fallen die Antworten der Schülerinnen und Schüler fachspezifisch aus. Es ließe sich vermuten, dass die Einstellungen des Schülerinnen und Schüler zum Fach das Antwortverhalten möglicherweise beeinflussen.

Angesichts verschiedener methodischer Überlegungen und Befunde (siehe Kap. 11) erscheint es zum einen angemessen, die Dimensionen in Form manifester Skalenmittelwerte zu modellieren und zum anderen spezifisch die Korrelationen innerhalb der einzelnen Dimensionen zu untersuchen. Dies wird zum einen mit der Heterogenität der Facetten innerhalb der fünf großen Dimensionen und der daraus resultierenden eher schwach ausgeprägten internen Konsistenz des BigFive-Modells entlang des Instruments des BFI-S (siehe Kap. 11.2) begründet. Zum anderen führt angesichts der hohen Modellkomplexität die Parameteranzahl eines latenten Modells zu nicht interpretierbaren Ergebnissen.¹³⁸

¹³⁸ In Bezug auf die latenten Gesamtmodelle sowie auf die latenten Einzelmodelle der jeweiligen Dimensionen zeigen sich zwar etwas höhere Korrelationen als in den manifesten Modellen, jedoch gibt Mplus außer in der Dimension *Offenheit* durchweg Fehlermeldungen aus, weshalb die Werte nicht zuverlässig sind. Bei freier Schätzung der Korrelationen

Forschungsfrage 1: Welcher Zusammenhang besteht zwischen selbst- und fremdwahrgenommener Lehrerpersönlichkeit?

Darüber hinaus unterscheiden sich die konstruktidentischen Skalen zu den Dimensionen *Verträglichkeit* und *Resilienz* auf latenter Ebene leicht zwischen den Schülerinnen/Schülern und ihren Lehrerinnen/Lehrern (siehe ebenfalls Kap. 11.2).

15.1 Zusammenhang von verschiedenen Wahrnehmungsperspektiven der Lehrerpersönlichkeit – Korrelationen

Um die Übereinstimmung der verschiedenen Perspektiven – jener der Schülerinnen und Schüler sowie jener der Lehrerinnen und Lehrer – hinsichtlich der Lehrerpersönlichkeit zu untersuchen, werden die Korrelationen zwischen den perspektivenspezifischen Einschätzungen untersucht. Es werden hier die Individualebene innerhalb der einzelnen Klassen (d.h. die individuellen Antworten werden um den Standardfehler korrigiert) als auch die Klassenebene, die Unterschiede zwischen den verschiedenen Klassen berücksichtigt, indem zum einen die *within*-Matrix, die die Korrelationskoeffizienten innerhalb der Klassen mittelt, dargestellt und darüber hinaus die *between*-Matrix, die Mittelwerte aus den Korrelationskoeffizienten, die zwischen den Klassen bestehen, generiert wird. Alle Berechnungen sind am Gesamtmittelwert zentriert. In einem ersten Schritt werden ausschließlich die Zusammenhangsmaße der beiden Perspektiven innerhalb der jeweiligen Dimensionen modelliert (siehe Tab. 24).

zwischen den Einzelitems je Skala resp. Dimension zeigt sich, dass ausschließlich innerhalb der Dimension *Extraversion* die einander entsprechenden Items signifikant miteinander zusammenhängen, darüber hinaus in den anderen Dimensionen jedoch die Korrelationen zwischen den Items der jeweiligen Skalen deutlich variieren und einander korrelativ nicht zuzuordnen sind. U.a. deshalb werden die Analysen im weiteren Verlauf dieser Arbeit auf Basis der manifesten Variablen durchgeführt.

Forschungsfrage 1: Welcher Zusammenhang besteht zwischen selbst- und fremdwahrgenommener Lehrerpersönlichkeit?

Tab. 24: Standardisierte Korrelationen zwischen den Lehrerselbsteinschätzungen und den Schülereinschätzungen (Fremdeinschätzung) der BigFive-Persönlichkeitsdimensionen (Deutsch: $N_{within}=1064/N_{between}=53$; Mathematik: $N_{within}=1042/N_{between}=50$; Signifikanz_{within}: $p<0,01$ /Signifikanz_{between}: $p<0,05$).

Individual-/Klassenebene	Deutsch	Mathematik
Gewissenhaftigkeit	,06/,16	,05/,11
Extraversion	,07/,12	,25**/,41**
Offenheit	,05/,12	,14/, 31*
Resilienz	,21**/,41**	,13/,24
Verträglichkeit	n.b.	,03/,09
	Fit-Indizes	
RMSEA	,09/,16	,08/,15
SRMR	,23/,30	,22/,32
CFI	,07/,01	,06/,01
TLI	-,08/-,12	-,06/-,08

Anmerkung: n.b. steht für nicht berechnet; signifikante Korrelationen sind fett markiert.

Die Zusammenhgangsrichtung ist bei allen Korrelationen sowohl innerhalb als auch zwischen den Klassen dieselbe (positiv), lediglich die Stärke ist verschieden ausgeprägt (was u.a. durch die unterschiedlichen Stichprobengrößen erklärbar ist). Es wird ersichtlich, dass die Schülerwahrnehmungen und die Selbstwahrnehmungen der Lehrerpersönlichkeit insgesamt eher schwach zusammenhängen, zumal die Fit-Werte (bis auf die RMSEA-Werte der individuellen *within*-Ebene) auf eine insgesamt nicht vorhandene Passung der beiden Wahrnehmungsperspektiven hinweisen.

Es zeigen sich dabei dimensionsspezifische Unterschiede zwischen den Fächern Deutsch und Mathematik: Während im Fach Deutsch die Schüler- und Lehrerwahrnehmungen in der Dimension *Resilienz* signifikant korrelieren, ist dies im

Forschungsfrage 1: Welcher Zusammenhang besteht zwischen selbst- und fremdwahrgenommener Lehrerpersönlichkeit?

Fach Mathematik in den Dimensionen *Extraversion* und etwas abgeschwächter *Offenheit für neue Erfahrungen* der Fall.

Abschließend wird noch ein Blick auf die Interkorrelationsmatrix zwischen Schüler- und Lehrereinschätzung über alle Persönlichkeitsdimensionen geworfen (Tab. 25). Damit wird geprüft, ob evtl. divergente Dimensionen stärker kovariieren und Schüler/Schülerinnen und Lehrer/Lehrerinnen möglicherweise unterschiedliche Auffassungen bzw. ein unterschiedliches Verständnis der jeweiligen gemessenen Dimensionen haben:

Tab. 25: Interkorrelationsmatrix der schülerperzipierten und der selbsteingeschätzten Lehrkraft-BigFive-Dimensionen auf Lehrkräftebene in Deutsch/Mathematik (N=53/50).

		<u>Deutsch/Mathematik</u>				
	Schüler	Gewissenhaftigkeit	Extraversion	Offenheit	Resilienz	Verträglichkeit
Gewissenhaftigkeit		,15/,11	-,06/,19	<u>,20/,20</u>	,08/,09	n.b./,04
Extraversion		,07/-,08	,10/,40**	,04/,15	<u>,12/,33*</u>	n.b./-,09
Offenheit		<u>,13/,09</u>	,09/,28	,11/,29*	,12/,26	n.b./-,09
Resilienz		,01/-,07	,09/,12	,07/,09	,36**/,23	n.b./-,02
Verträglichkeit		,07/,07	<u>,12/-,02</u>	<u>,07/,11</u>	<u>,13/,07</u>	n.b./,09

Anmerkung: n.b. steht für nicht berechnet; Hervorhebungen: **gerahmt** – Korrelationen innerhalb der entsprechenden Dimensionen; unterstrichen – höchste Korrelationen außerhalb derselben Dimension; **fett** – signifikante Korrelationen.

Die Interkorrelationsmatrix zeigt, dass im Fach Deutsch innerhalb der Dimension *Resilienz* und im Fach Mathematik innerhalb der Dimensionen *Extraversion* und *Offenheit* weiterhin die stärksten Zusammenhänge zwischen den Schüler- und Lehrerperspektiven bestehen. Die weiteren Dimensionen korrelieren stärker mit anderen Dimensionen als innerhalb der ihren. So hängt im Fach Mathematik bspw. die von den Lehrerinnen und Lehrern selbsteingeschätzte *Resilienz* stärker mit der von den Schülerinnen und Schülern eingeschätzten *Extraversion* zusammen als mit der von den Schülerinnen und Schülern eingeschätzten *Resilienz*. Insgesamt lassen sich relativ hohe kongruente Korrelationen beobachten, jedoch

Forschungsfrage 1: Welcher Zusammenhang besteht zwischen selbst- und fremdwahrgenommener Lehrerpersönlichkeit?

ändert die Betrachtung nichts daran, dass die beiden Perspektiven nicht in nennenswerter Weise miteinander korrelieren – bis auf die Dimensionen *Resilienz* im Fach Deutsch sowie knapp bzw. vor allem auf Individualebene die Dimension *Extraversion* im Fach Mathematik.

15.2 Kontrolle des Lehrkraftgeschlechts im Fach Mathematik

Die unterschiedlichen Zusammenhangsmuster gehen möglicherweise auf ungleiche Anteile an weiblichen und männlichen Lehrkräften in den beiden betrachteten Fächern zurück: in Mathematik unterrichten deutlich mehr männliche als weibliche Lehrkräfte. Zudem wurden zuvor (vgl. Kap. 14.1.1 und 14.2.3) signifikante Differenzen zwischen männlichen und weiblichen Lehrkräften in der Dimension *Extraversion* seitens der Selbsteinschätzungen der Lehrkräfte und in der Dimension *Resilienz* (sowie sehr knapp verfehlt in der Dimension *Offenheit*) seitens der Schülereinschätzungen identifiziert. Daher werden im folgenden Schritt die Korrelationen im Fach Mathematik erneut unter Kontrolle des Geschlechts der Lehrkraft berechnet. Das Geschlecht der Lehrkraft als potentielles Bias wurde dabei als Moderatorvariable in die Analysen aufgenommen und somit wurden die Korrelationen zwischen den Schüler- und Lehrereinschätzungen um dieses bereinigt.

Tab. 26: Standardisierte Korrelationskoeffizienten von Schülerfremdeinschätzung der Lehrerpersönlichkeit mit Lehrkräfteselbsteinschätzung im Fach Mathematik unter Einbezug des Lehrkraftgeschlechts als Moderatorvariable ($N_{within}=1042$, $N_{between}=50$; Signifikanzniveau auf Individual-/Klassenebene: $p<0,01/<0,05$).

Mathematik	Gewissenhaftigkeit	Extraversion	Offenheit	Resilienz	Verträglichkeit
<i>within/between</i>	,06/,11	,29**/,47**	,14/,32*	,11/,23	,04/,09

Anmerkung: Da sich die Modellgüteindices nicht nennenswert gegenüber den o.g. verändern, werden diese hier nicht erneut berichtet.

Bei Kontrolle des Lehrkraftgeschlechts im Fach Mathematik bleiben die Zusammenhangsmaße der Schüler- und Lehrereinschätzungen innerhalb der Dimensi-

Forschungsfrage 1: Welcher Zusammenhang besteht zwischen selbst- und fremdwahrgenommener Lehrerpersönlichkeit?

onen *Resilienz* und *Offenheit* nahezu unverändert. Innerhalb der Dimension *Extraversion* tritt die bereits signifikante Ausprägung etwas weiter hervor. Angesichts der Ergebnisse, die sich unter dem Einbezug des Geschlechts der Lehrkraft eröffnen (siehe diverse Korrelationsmatrizen in Anhang D), lässt sich schließen, dass diese einer ausführlichen separaten und präzisen Analyse bedürfen. Da Geschlechterfragen jedoch nicht im Fokus dieser Arbeit stehen, wird dieser Frage im Rahmen dieser Arbeit nicht weiter nachgegangen. Festhalten lässt sich hier lediglich, dass das Geschlecht der Lehrkraft die Korrelation zwischen Lehrer- und Schülereinschätzungen kaum zu beeinflussen scheint, allerdings je nach Dimension und möglicherweise auch je nach Fach unterschiedlich.

16 Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten *Lehrerpersönlichkeit* und *Unterrichtsqualität*?

Die zweite Forschungsfrage, gleichwohl Hauptaugenmerk dieser Arbeit, richtet sich auf die Beziehung zwischen Lehrerpersönlichkeit und Unterrichtsqualität. Gemäß den theoretischen Erkenntnissen und entsprechenden Forschungsbefunden (vgl. Kap. 2) wird im Unterschied zum Vorgehen zur Beantwortung der ersten Forschungsfrage, die sich auf ungerichtete Zusammenhänge (Korrelationen) beschränkt, hier von einer Wirkungsrichtung der Konstrukte ausgegangen. Demnach lautet die Hypothese, dass sich die Lehrerpersönlichkeit auf die Wahrnehmung der Unterrichtsqualität auswirkt; nicht per se als solche, allerdings durchaus innerhalb derselben Perspektive. Das heißt konkret: Nicht die Selbsteinschätzung der Persönlichkeit durch die Lehrkraft steht unter der Annahme, sich auf die von den Schülerinnen und Schülern wahrgenommene Unterrichtsqualität auszuwirken,¹³⁹ sondern die Einschätzungen der Schülerinnen und Schüler der Lehrerpersönlichkeit auf deren Einschätzung von Unterrichtsqualität. Wie anhand der Analysen bezüglich der Forschungsfrage 1 zum Zusammenhang zwischen selbst- und fremdwahrgenommener Lehrerpersönlichkeit gezeigt werden konnte, bestehen kaum signifikante Korrelationen zwischen den Wahrnehmungen der Schülerinnen und Schüler und den Selbstwahrnehmungen der Lehrkräfte bezüglich der Lehrerpersönlichkeit. Auch angesichts dessen erscheint es kaum angemessen, hinsichtlich der zweiten Forschungsfrage auf die Persönlichkeits-selbsteinschätzungen der Lehrerinnen und Lehrer näher einzugehen. Dennoch werden die korrelativen Befunde diesbezüglich vollständigkeithalber im folgenden Kapitel berichtet.

¹³⁹ Dieser Ansatz entspricht dem in der Forschung gängigen Vorgehen, wenn *die* (vermeintlich als manifester Kern angenommene) Lehrerpersönlichkeit in Zusammenhang mit Unterrichtsqualität gebracht wird (siehe hierzu Kap. 2).

16.1 Perspektivenübergreifender Zusammenhang von Persönlichkeitsselbsteinschätzungen der Lehrkräfte mit schülerperzipierten Prozessmerkmalen des Unterrichts – Korrelationen und Interpretation

Um einen Einblick in die Daten bezüglich des bislang in der Unterrichtsfor-
schung üblichen Ansatzes der Kombination von Lehrer- und Schülerperspekti-
ven zu erhalten, werden an dieser Stelle die Persönlichkeitsselbsteinschätzungen
der Lehrkräfte mit den Schülerwahrnehmungen der Unterrichtsqualität in Zu-
sammenhang gebracht.¹⁴⁰

Zwischen den Selbsteinschätzungen der Lehrerpersönlichkeit und den aggregier-
ten Einschätzungen des entsprechenden Unterrichts durch die Schülerinnen und
Schüler zeigen sich nahezu keine signifikanten Korrelationen. Lediglich hin-
sichtlich der Prozessmerkmale *Störungsprävention* und *angemessenes Interakti-
onstempo* zeigen sich einzelne moderate Zusammenhänge: Im Fach Mathematik
korreliert die *Störungsprävention* mit 0,35 signifikant mit der *Extraversion*, die
jedoch mit einer Kovarianz des Lehrkraftgeschlechts von 0,19 einhergeht. Im
Fach Deutsch korreliert die selbsteingeschätzte *Lehrerresilienz* mit der schüler-
perzipierten *Angemessenheit des Interaktionstempos* zwar signifikant, jedoch
nur knapp moderat (mit 0,29). *Gewissenhaftigkeit* und *Störungsprävention* ver-
fehlen im Fach Deutsch knapp die Signifikanzgrenze.

¹⁴⁰ Die Berechnung erfolgt mit SPSS 24, die Schülerdaten wurden auf Lehrkraftebene ag-
gregiert.

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersönlichkeit und Unterrichtsqualität?

Tab. 27: Korrelationen der BigFive-Selbsteinschätzungen der Lehrerinnen und Lehrer in Deutsch/Mathematik mit den schülerperzipierten Prozessmerkmalen des Unterrichts; manifest und auf Lehrkraftebene aggregiert ($N_{\text{Deutsch}}=53/N_{\text{Mathematik}}=50$; Signifikanz auf 5%-Niveau).

Deutsch/Mathematik	Gewissenhaftigkeit	Extra-version	Offenheit	Resilienz	Verträglichkeit
Methodenvariation	-,04/,01	-,01/,24	,20/,19	,18/,12	,15/-,09
Motivierung	,02/-,11	-,05/,22	-,03/,02	,17/,07	,04/-,24
Interaktionstempo	-,05/-,09	,13/,07	-,06/,12	,29**/,07	,00/,00
Alltagsbezug	,14/-,12	,03/,12	-,09/,06	,09/,03	-,06/,24
Erklärkompetenz	,09/-,05	-,01/,15	,08/,15	,17/,02	,04/-,09
Störungsprävention	,27 ⁺ /-,12	-,07/,35**	,09/,09	-,04/,18	,05/-,22
Disziplin	-,02/-,11	,01/,23	,06/-,16	,14/-,02	-,18/-,19

Diese Befunde stehen im Widerspruch zu jenen von Othmann (2009, siehe Kap. 2.2), die signifikante Korrelationen der (selbsteingeschätzten) Persönlichkeitsdimensionen *Extraversion*, *Verträglichkeit* und *Gewissenhaftigkeit* mit der Lehreffektivität nachweist. Allerdings ist zu beachten, dass global erhobene Einschätzungen der Lehreffektivität zum einen ein anderes Konstrukt abbilden und zum anderen generell stärker biasgefährdet sind als die in der vorliegenden Untersuchung differenzierten Prozessmerkmale des Unterrichts – obgleich diese, wie in diversen Studien nachgewiesen wurde (vgl. Kap. 4.4.1 und 2.2.2), durchaus auch auf die Lehreffektivität abzielen. Othmann (2009) konnte zudem keine Beziehungen zwischen Lehreffektivität und *Neurotizismus* und *Offenheit* ausmachen, während in den vorliegenden korrelativen Befunden im Fach Deutsch die Dimension *Resilienz* (also der Gegenpol von *Neurotizismus*) die einzige ist, die eine signifikante Korrelation und zwar mit dem Unterrichtsmerkmal *angemessenes Interaktionstempo* aufweist.

Eine weitere Studie untersuchte die Korrelationen von BigFive-Persönlichkeits-selbsteinschätzungen mit archivierten Studentenevaluationen zur Unterrichtseffektivität (Kneipp/Kelly et al. 2010). Dabei korrelierte lediglich die Dimension

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersönlichkeit und Unterrichtsqualität?

Verträglichkeit signifikant mit den Studenturteilen und erwies sich – im Gegensatz zu den anderen Dimensionen – als stärkster Prädiktor. Auch diesem Befund widersprechen die Korrelationen der vorliegenden Untersuchung: Es lassen sich keine nennenswerten Korrelationen zwischen der selbsteingeschätzten *Verträglichkeit* der Lehrkräfte mit einem der schülerperzipierten Prozessmerkmale des Unterrichts ausmachen.

Feldmann (1986) hingegen generierte eine Sammlung von persönlichen Merkmalen unabhängig von den BigFive, denen er allgemeine, globale Effektivitätseinschätzungen des Unterrichts (*teacher effectiveness in the classroom*, *instructional effectiveness* und *teacher effectiveness*) gegenüberstellte. Er kommt dabei zu dem Ergebnis, dass Persönlichkeitseinschätzungen von Lehrerinnen und Lehrern mit den Schülereinschätzungen der Unterrichtseffektivität eher in wenigen Fällen und zudem schwach bis sehr schwach ($<0,30$) korrelieren. Ebenso kommen Hart und Driver 1978 in ihrer Untersuchung zu dem Befund, dass weder Ähnlichkeiten zwischen dem Persönlichkeitstypus der Lehrkraft und dem des Lernenden zur Wahrnehmung höherer Lehreffektivität führen noch anderweitige Zusammenhänge zwischen Lehrer- und Schülerwahrnehmungen bestehen.

Baier et al. (2018) weisen in Ihrer Studie zwar ebenfalls einen perspektivenübergreifenden Effekt von Extraversion auf die Unterrichtsqualität im Fach Mathematik nach. Allerdings bezieht sich dieser Effekt zwar auf lernunterstützendes Verhalten sowie die kognitive Aktivierung im Unterricht, nicht aber auf Aspekte der Störungsprävention. Die Störungsprävention scheint der Studie zufolge negativ von der Gewissenhaftigkeit beeinflusst, was sich allerdings mit dem hier vorgefundenem Befund nicht bestätigen lässt.

Angesichts der Befunde von Feldman (1986), Hart und Driver (1978), Baier et al. (2018) und jenen von Othmann (2009), die nachwies, dass sich trotz hoher Korrelationen kaum perspektivenübergreifende Wirkungseffekte von Persönlichkeitseigenschaften auf die Lehreffektivität zeigen, wird aus den äußerst schwachen Korrelationen der vorliegenden Untersuchung die Konsequenz gezogen, die Konstrukte der *Lehrerpersönlichkeit* und der *Unterrichtsqualität* nicht perspektivenübergreifend zu verknüpfen. Der Zusammenhang von der aus Lehrerperspektive erfassten Selbsteinschätzung der Persönlichkeit und der aus Schülerperspektive erfassten Unterrichtsqualität wird daher nicht in weiteren Analysen verfolgt.

16.2 Perspektivenspezifischer Zusammenhang von Lehrerpersönlichkeit und Unterrichtsqualität – Regressionsanalysen

Die hinsichtlich der Schülerinnen- und Schülerwahrnehmungen perspektivenspezifischen Regressionsmodelle werden gemäß den Konstruktvalidierungen (vgl. u.a. Kap. 11.3.2 und 12.3.1) fachspezifisch in Deutsch und Mathematik durchgeführt. Es wurde deutlich, dass die Schülerurteile zwischen den Fächern unterschiedlich ausfallen und es demnach nicht legitim erscheint, diese Differenzen durch die Analyse fachübergreifender Modelle zu ignorieren. Wie bei Forschungsfrage 1 erscheint es auch an dieser Stelle angemessen, die latenten Konstrukte der Persönlichkeitsdimensionen als auch der Unterrichtsmerkmale in Form manifester Skalenmittelwerte zu modellieren (siehe hierzu Kap. 10, 11.1, 11.2).¹⁴¹ In diesem Sinne wird untersucht, welchen Einfluss die von den Schülerinnen und Schülern wahrgenommene Lehrerpersönlichkeit auf die Schülerurteile zur Unterrichtsqualität im Fach Mathematik sowie vergleichend im Fach Deutsch hat. Dazu wurden in einem ersten saturierten Modell (Modell 1a) Mehrebenen-Strukturmodelle entlang einer multiplen Regression der differenzierten Prozessmerkmale des Unterrichts auf die Persönlichkeitsdimensionen *Gewissenhaftigkeit*, *Extraversion*, *Offenheit für neue Erfahrungen*, *Resilienz* und *Verträglichkeit* erstellt, um sowohl individuelle als auch Klasseneffekte zu untersuchen und gegenüberzustellen (Modellstruktur siehe Abb. 13, Kap. 13.2, Ergebnisse sind Anhang E zu entnehmen). In einem weiteren, spezifizierten Modell (Modell 1b) werden dann ausschließlich die im saturierten Modell ermittelten signifikanten Pfade aufgenommen.

¹⁴¹ Ein latentes Mehrebenenmodell, das alle Parameter beider Konstrukte gleichzeitig einbezieht, ist selbst für Mplus zu komplex. Allerdings zeigt sich bei einer skalenzentrierten (gegenüber der in der Arbeit gewählten konstruktorientierten) Analyse in beiden Fächern gleichermaßen, dass bei separater und latenter Betrachtung der einzelnen Persönlichkeitsdimensionen auf individueller Ebene innerhalb der Klassen alle (einzelnen) Persönlichkeitsdimensionen signifikante Effekte auf die Wahrnehmung der Unterrichtsqualität haben. Die Signifikanz der Effekte verschwindet jedoch gänzlich auf der Ebene zwischen den Klassen, d.h. auf Klassenebene bestehen bei separater, skalenorientierter Betrachtung der von den Schülerinnen und Schülern an ihren Lehrerinnen und Lehrern wahrgenommenen Persönlichkeitsdimensionen keinerlei Wirkungen auf die wahrgenommenen Unterrichtsqualität.

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersönlichkeit und Unterrichtsqualität?

Da vor allem die Dimensionen *Extraversion* und *Offenheit* (insbesondere auf Klassenebene) sehr hoch korreliert sind (siehe Kap. 14.2.4) und Mplus in multiplen Regressionsanalysen per Voreinstellung Korrelationen zwischen den Prädiktoren herausrechnet, besteht die Gefahr der Multikollinearität und damit einhergehenden verzerrten Regressionskoeffizienten (β -Gewichte). Von einer nennenswerten Multikollinearität der Prädiktoren kann zwar insofern Abstand genommen werden, als diese nicht perfekt oder mit $r > 0,9$ (siehe Kap. 14.2.4, vgl. Wooldridge 2013, Allison 2012) korreliert¹⁴² sind und zudem auf latenter Ebene keinen gemeinsamen Faktor bilden (siehe Kap. 11.3). Allerdings verfehlt die Korrelation der Dimensionen *Extraversion* und *Offenheit* die Ausprägung der Toleranzgrenze nur knapp. Zudem muss von einem essentiellen, d.h. inhaltlich bedingten Zusammenhang der beiden Dimensionen ausgegangen werden, da diese generell hoch korrelieren (vgl. Kap. 3.3.2, 14.2.4). Daher wird untersucht, wie sich die Effektstärken innerhalb eines modifizierten Strukturgleichungsmodells abbilden: Dieses Modell (Modell 2a) entspricht dem saturierten Modell, in das zudem die Korrelation zwischen *Extraversion* und *Offenheit* aufgenommen wird (siehe hierzu Abb. 14, Kap. 13.2). Äquivalent zum Vorgehen im ersten Modell werden daraufhin die signifikanten Pfade und zusätzlich die Korrelation zwischen *Extraversion* und *Offenheit* in ein weiteres Modell aufgenommen (Modell 2b). Abschließend wird ein gemeinsamer Faktor (Ladungen siehe Anhang F) über die BigFive-Dimensionen gebildet und eine einfache lineare Regression durchgeführt (Modellstruktur siehe Abb. 15, Kap. 13.2); zum einen, um einen Überblick über den Effekt der Persönlichkeit als Gesamtkonstrukt zu erhalten (Modell 3) und zum anderen, um das Ausmaß durch Multikollinearität verzerrter Varianzaufklärung und Regressionskoeffizienten präziser abschätzen zu können.¹⁴³ Die Model-Fit-Indizes sind der folgenden Tabelle zu entnehmen:

¹⁴² Eine perfekte Kollinearität liegt vor, wenn zwei Vektoren parallel oder übereinstimmend verlaufen (mit $r=1$ bzw. -1). Um das Kriterium der Kollinearität zu erfüllen, wird in der aktuellen Literatur ein Zusammenhangsmaß von $r=0,9$ angesetzt (vgl. Allison 2012, Stapleton 2016, Muthén 2017).

¹⁴³ Die Ergebnisse eines Regressionsmodells mit Faktoren 2. Ordnung auf beiden Konstrukten (Persönlichkeit und Unterrichtsqualität) findet sich vollständigshalber in Anhang F.

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersönlichkeit und Unterrichtsqualität?

Tab. 28: Güteindizes der Strukturmodelle in Deutsch und Mathematik ($N_{within}= 2247/$
 $N_{between}=123$).

Modell	Fach	Chi ² /df	CFI	TLI	RMSEA	SRMR _w / SRMR _b	AIC/ BIC
1a saturiert	<u>Deutsch</u> Mathematik						
							saturiert
1b sig. Pfade	<u>Deutsch</u> Mathematik	52,14/18 126,27/33	,99 ,98	,96 ,93	,03 ,04	,02 _w /,12 _b ,03 _w /,08 _b	48609,29/ 49146,26 61906,48/ 62477,73
2a saturiert + Korr. E/O	<u>Deutsch</u> Mathematik	646,16/12 587,31/12	,88 ,89	-,22 -,19	,15 ,15	,12 _w /,25 _b 11 _w /,28 _b	62783,76/ 63589,21 64721,21/ 65526,67
2b sig. Pfade + Korr. E/O	<u>Deutsch</u> Mathematik	1267,60/50 1325,72/46	,78 ,77	,44 ,36	,10 ,11	,14 _w /,30 _b ,14 _w /,27 _b	63304,74/ 63893,12 63352,57/ 63963,80
3 BigFive als einz. Faktor	<u>Deutsch</u> Mathematik	503,30/66 589,03/66	,93 ,91	,86 ,83	,05 ,06	,03 _w /,10 _b ,04 _w /,09 _b	62479,07/ 63062,24 64739,40/ 65322,66

Vor allem im Fach Deutsch (aber auch im Fach Mathematik) wird offensichtlich, dass das Modell (1b) mit signifikanten Pfaden und ohne Korrelationen zwischen *Extraversion* und *Offenheit* die Daten am besten abbildet. Die Werte von Modell 2 (a und b) weisen eher auf eine Fehlspezifikation hin. Daher kann dieses Modell verworfen werden und wird im weiteren Verlauf nicht berücksichtigt (Effekte und Varianzaufklärungen lediglich in Anhang E). Das übergeordnete lineare Modell 3 wiederum hebt den Konstruktcharakter der BigFive-Dimensionen durchaus deutlich hervor und kann zur Verifizierung von Modell 1 (a und b) herangezogen werden.

Die folgende Tabelle zeigt die Regressionskoeffizienten und Anteile erklärter Varianz des spezifizierten Modells 1b:

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersonlichkeit und Unterrichtsqualität?

Tab. 29: Modell 1b – signifikante Effekte (Regressionskoeffizienten) der Lehrerpersonlichkeit auf die Unterrichtsqualität.

Fach	Prädiktoren	Disziplin	Störungsprävention	Erklärkompetenz	Alltagsbezug	Interaktionstempo	Motivierung	Methodenvariation
Deutsch	Gewissenhaftigkeit	,14**	,20**	,25**	,10**		,15**	,11**
	Extraversion							
	Offenheit		,15**	,19**	,26**		,33**	,23**
	Resilienz	,08**						
Mathematik	Verträglichkeit	,08**		,20**	,09**	,32**	,17**	,12**
	Erklärte Varianz (R ²)	,06**	,09**	,26**	,13**	,10**	,27**	,13**
	Gewissenhaftigkeit	,13**	,17**	,21**			,08**	
	Extraversion							
Deutsch	Offenheit		,13**	,17**	,25**		,29**	,23**
	Resilienz	,10**		,20**	,12**	,29**	,18**	,15**
	Verträglichkeit	,04**	,06**	,21**	,10**	,09**	,20**	,10**
	Erklärte Varianz (R ²)	,35**	,64**	,45**				,23**
Mathematik	Gewissenhaftigkeit	,49**		,39**	,36**	,38**	,56**	
	Extraversion							
	Offenheit							
	Resilienz	-,34**	-,52**	,18**		,72**		
Deutsch	Verträglichkeit	,48**	,56**	,60**	,13[†]	,82**	,31*	n.s.
	Erklärte Varianz (R ²)	,49**	,82**	,58**	,40**		,25**	
	Gewissenhaftigkeit							
	Extraversion		,21					
Mathematik	Offenheit				,30**		,44**	,53**
	Resilienz	,57**		,23**			,26**	,30**
	Verträglichkeit	-,42**	-,60**			,85**		
	Erklärte Varianz (R ²)	,61**	,70**	,78**	,40**	,73**	,74**	,36**

Anmerkung: Moderate und starke Effekte – gerundet $\geq 0,30$ – sind fett gekennzeichnet; n.s. steht für nicht signifikant. Individualebene: $N_{within}=2247/Signifikanz_{within}=p<0,01$; Klassenebene: $N_{between}=123/Signifikanz_{between}=p<0,05$.

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersönlichkeit und Unterrichtsqualität?

Hinsichtlich der *Methodenvariation* – also dem je nach Fach und Lerninhalt adäquaten Einsatz diverser Unterrichtsmethoden – zeigen sich auf Klassenebene die deutlichsten Fachunterschiede zwischen Deutsch und Mathematik; in Modell 1 (a und b) ebenso wie in Modell 3 (siehe Tab. 30), was auf verlässliche Werte v.a. des saturierten Modells schließen lässt. Eine mittlere Varianzaufklärung ist bei den Prozessmerkmalen *Alltagsbezug*, *Störungsprävention* und *Disziplin* zu verzeichnen.

Tab. 30: Modell 3 – standardisierte Regressionskoeffizienten und Varianzaufklärung der Prozessmerkmale des Unterrichts durch einen übergeordneten Persönlichkeitsfaktor 2. Ordnung ($N_{within} = 2247/N_{between} = 123$).

within/between	Deutsch	R ²	Mathematik	R ²
Disziplin	,26/,63	,07/,40	,27/,64	,07/,41
Störungsprävention	,30/,53	,09/,28	,33/,42	,11/,18
Erklärkompetenz	,54/,92	,30/,85	,53/,89	,28/,79
Alltagsbezug	,35/,70	,12/,49	,31/,72	,10/,51
Interaktionstempo	,28/,70	,08/,49	,38/,78	,15/,60
Motivierungsqualität	,51/,94	,26/,89	,48/,93	,23/,86
Methodenvariation	,36/,67	,13/,45	,24/,84	,06/,71

Anmerkung: Alle Effekte sind signifikant auf 1%-Niveau (**).

Aus den Anteilen aufgeklärter Varianz der *Störungsprävention*, die in den differenzierten Modellen 1 und 2 deutlich höher liegt als im linearen Regressionsmodell mit übergeordnetem Faktor (Modell 3), lässt sich ableiten, dass auch hier eine multikollinearitätsbedingte Verzerrung über die Dimensionen *Extraversion* und *Offenheit* hinaus besteht. In Modell 2 wird ersichtlich, dass die Modellierung der Korrelation zwischen *Extraversion* und *Offenheit* die Varianzaufklärung des *Interaktionstempos* (siehe Anhang E) eher steigert, was – ebenso wie die Güteindizes – auf eine Fehlspezifikation hindeutet. Insbesondere scheinen die Effekte auf das *Interaktionstempo* sowie dessen aufgeklärte Varianz durch die essentielle, konstruktimmanente Multikollinearität der Prädiktoren beeinträchtigt.

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersönlichkeit und Unterrichtsqualität?

Durch die ausschließliche Aufnahme der signifikanten Pfade werden implizit andere Prädiktoren ausgeblendet und somit die Gefahr der Multikollinearität herabgesetzt (Modelle 1b und 2b). Dadurch können verzerrte Varianzanteile und Regressionseffekte (teilweise) korrigiert und somit interpretiert werden. Jedoch verbessert dies die Modellgüteindizes von Modell 2b nicht. Daher werden diese Werte nicht weiter berücksichtigt und interpretiert, sondern die Befunde entlang der Modelle 1 und 3 beschrieben.

Angesichts der Modellgüteindizes sowie der dargestellten Befunde hinsichtlich der Effekte und Varianzanteile der verschiedenen Modelle scheinen die Modelle 1a und b die Daten adäquat widerzuspiegeln. Im Folgenden werden die Ergebnisse fachspezifisch für die Fächer Deutsch und Mathematik abgebildet und beschrieben, dabei stehen v.a. die Werte aus Modell 1b im Vordergrund.

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersönlichkeit und Unterrichtsqualität?

16.2.1 ...im Fach Deutsch

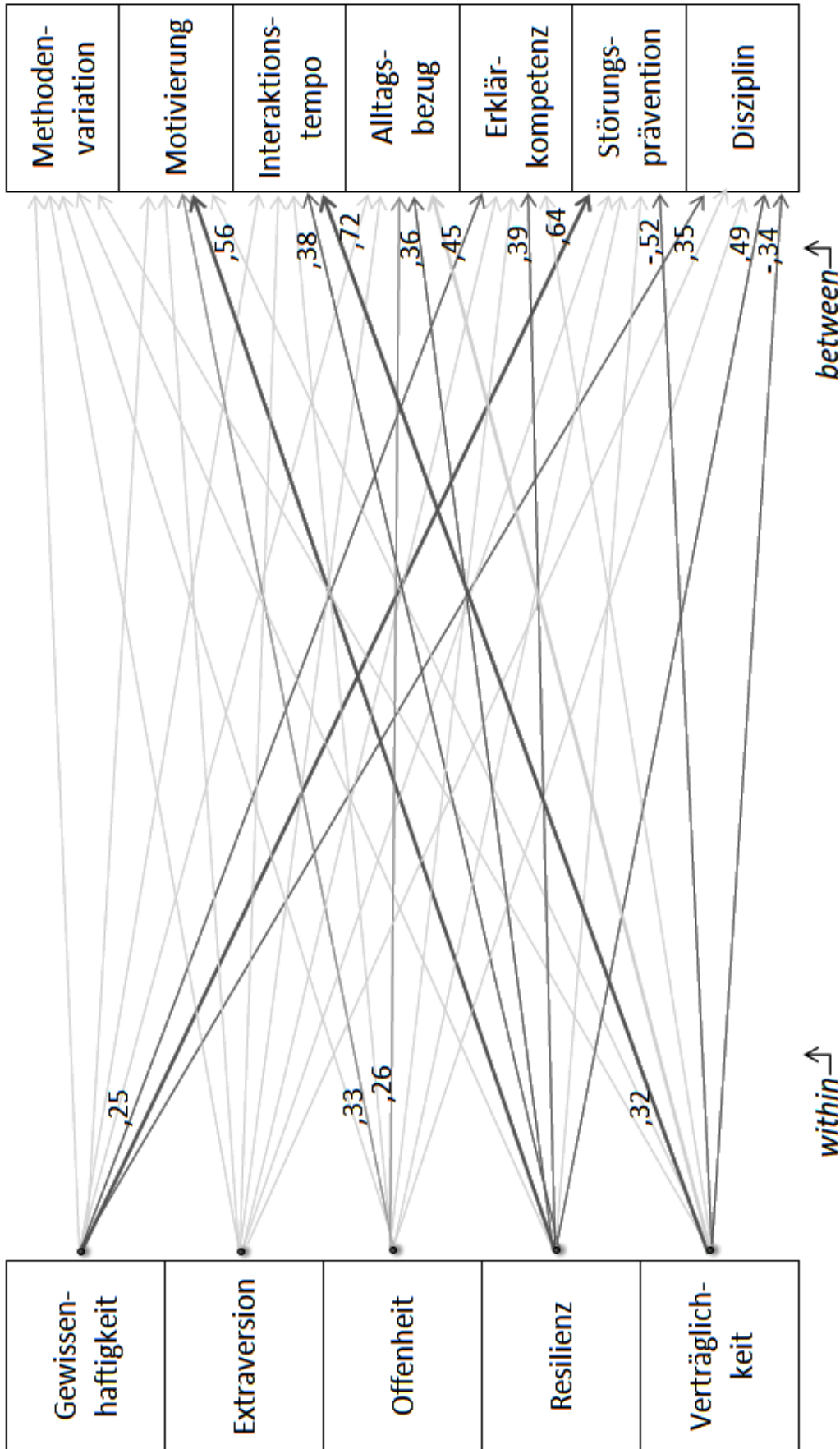


Abb. 27: Überblick über Modell 1b im Fach Deutsch – signifikante ($p > ,01$) und moderat bis stark ausgeprägte (gerundet $\geq ,30$ bzw. $\geq ,60$) Effekte von Schülereinschätzungen der Lehrerpersönlichkeit auf die Wahrnehmung der Unterrichtsqualität auf Individual- und auf Klassen- resp. Lehrkräfteebene ($N_{within} = 2247$ und $N_{between} = 123$).

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersönlichkeit und Unterrichtsqualität?

Auf Individualebene der Schülerinnen und Schüler (*within*¹⁴⁴) lassen sich im Fach Deutsch folgende Effekte in Modell 1b beobachten: Die Dimension *Extraversion* lässt keinen und die *Resilienz* lediglich einen signifikanten, zumal sehr schwach ausgeprägten Effekt auf Prozessmerkmale der Unterrichtsqualität verzeichnen. Dahingegen zeigen die drei übrigen Dimensionen *Gewissenhaftigkeit*, *Offenheit* und *Verträglichkeit* auf fünf (*Offenheit*) bzw. sechs (*Gewissenhaftigkeit* und *Verträglichkeit*) der sieben verschiedenen Unterrichtsmerkmale signifikante Effekte. Angesichts ihrer moderat ausgeprägten β -Gewichte sind vor allem die Effekte der *Verträglichkeit* auf das *angemessene Interaktionstempo* und die *Offenheit* auf die *Motivierungsqualität* von nennenswerter Bedeutung. Darüber hinaus gilt es aufgrund des verhältnismäßig hohen Anteils an erklärter Varianz für die Prozessmerkmale *Erklärkompetenz* und *Motivierung* zudem auf die Dimensionen *Gewissenhaftigkeit* und *Verträglichkeit* zu verweisen, die für dieses Prozessmerkmal die stärksten Prädiktoren bilden.

Die Regressionskoeffizienten sowie die Varianzaufklärung unterscheiden sich kaum zwischen Modell 1a und Modell 1b, die Modifikation verstärkt die Tendenzen des saturierten Modells lediglich leicht. Die Varianzaufklärung des 3. Modells, in dem die Prozessmerkmale auf einen Persönlichkeitsfaktor 2. Ordnung regrediert werden unterstützen zudem die Werte aus Modell 1 dahingehend, als sie in ihren Tendenzen vergleichbar ausgeprägt sind.

Die Effekte der Dimension *Verträglichkeit* bleiben auf Klassen- resp. Lehrkräfteebene (*between*¹⁴⁵) in den Modellen 1a und 1b teilweise bestehen. Allerdings zeigen sie sich hinsichtlich der Unterrichtsmerkmale *Erklärkompetenz*, *Störungsprävention* und *Disziplin* in entgegengesetzter, negativer Wirkungsweise. D.h. auf Individualebene positive Zusammenhänge – höhere *Verträglichkeit* bedingt die Einschätzung höherer Unterrichtsqualität – verkehren sich bei den genannten Unterrichtsmerkmalen auf der Klassenebene ins Gegenteil: Insgesamt weniger verträglich wahrgenommene Lehrkräfte werden dort in ihrer Unterrichtsqualität positiver eingeschätzt. Mit der Modifizierung (Modell 1b) schwindet dieser Effekt lediglich in Bezug auf die *Erklärkompetenz*. Zudem ist auch auf Klassenebene – ebenso wie auf der Individualebene – die *Verträglichkeit* jene

¹⁴⁴ In Bezug auf einen konkreten Lehrer bzw. eine konkrete Lehrerin lassen sich innerhalb der einzelnen Klassen Regressionseffekte verzeichnen.

¹⁴⁵ Diese Effekte unterscheiden sich signifikant zwischen den Klassen.

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersönlichkeit und Unterrichtsqualität?

Dimension, die sich am deutlichsten auf das *Interaktionstempo* auswirkt. Insgesamt vier der auf Individualebene sechs regredierte Prozessmerkmale bleiben auch auf Klassenebene signifikant beeinflusst vom Ausmaß der schülerperzipierten *Gewissenhaftigkeit* der Lehrkraft; drei davon weisen moderate bis starke Regressionskoeffizienten auf. Die Dimensionen *Extraversion* und *Offenheit* lassen auf Klassenebene keinerlei signifikante Effekte verzeichnen. Bemerkenswert ist die Differenz zwischen Klassenebene und Individualebene in der Dimension *Resilienz*: Während sich im Prinzip keine nennenswerten Effekte auf Individual-ebene ausmachen ließen, sind fünf Prozessmerkmale auf Klassenebene signifikant beeinflusst von der insgesamt wahrgenommenen *Resilienz* der Lehrkraft, d.h. je psychisch stabiler eine Lehrkraft im Fach Deutsch erscheint, desto besser wird auch ihre Unterrichtsqualität eingeschätzt.

Insgesamt weichen die Effekte zwischen saturiertem und modifiziertem Modell eher moderat voneinander ab (bis auf die Regression der *Motivierung* auf die *Resilienz* und jene der *Erklärkompetenz* auf die *Verträglichkeit*). Die Werte der erklärten Varianz hingegen sinken in Modell 1b (gegenüber dem saturierten Modell 1a) deutlich ab – v.a. hinsichtlich der *Methodenvariation*, der *Motivierung* und des *Alltagsbezugs*. Die in Modell 3 aufgeklärte Varianz der Prozessmerkmale unterstützt zwar weitgehend das saturierte Modell (*Erklärkompetenz*, *Alltagsbezug*, *Motivierung* und *Methodenvariation* verfügen über äquivalente Werte, während das R^2 der *Disziplin* eher Modell 1b entspricht). Die Varianzaufklärungen des *Interaktionstempos* (87%) sowie der *Störungsprävention* variieren hingegen stark zwischen den 1. Modellen und Modell 3, was auf möglicherweise verzerrte Werte aufgrund der Multikollinearität hindeutet. Diese Werte sind somit vorsichtig zu interpretieren und verstärkt auf ihre theoretische Nachvollziehbarkeit zu hinterfragen.

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersonlichkeit und Unterrichtsqualität?

16.2.2 ...im Fach Mathematik

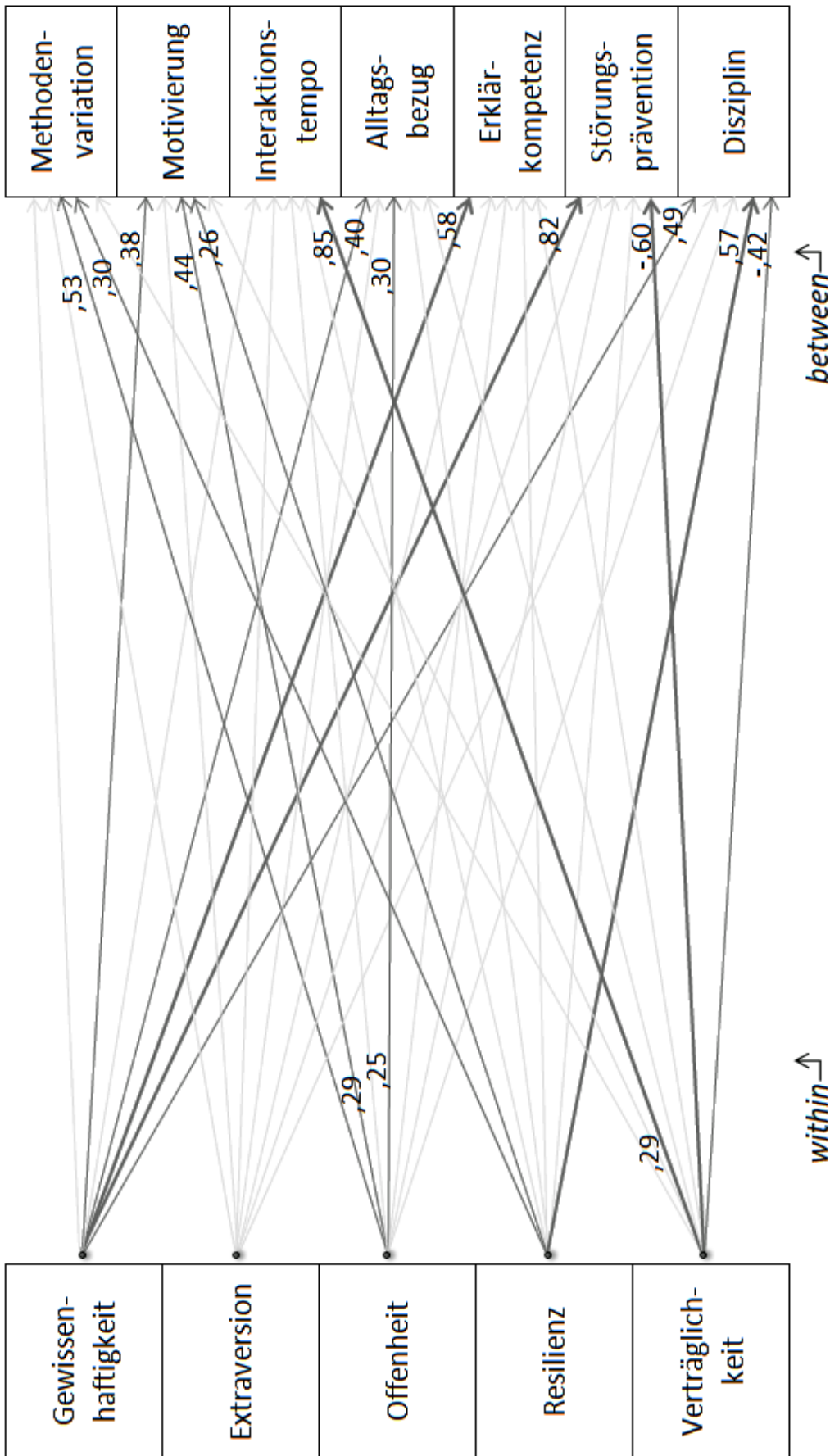


Abb. 28: Überblick über Modell 1b im Fach Mathematik – signifikante ($p > ,01$) und moderat bis stark ausgeprägte (gerundet $\geq ,30$ bzw. $\geq ,60$) Effekte von Schilereinschätzungen der Lehrerpersonlichkeit auf die Wahrnehmung der Unterrichtsqualität auf Individualebene ($N_{within}=2249$) und auf Klassen resp. Lehrkräfteebene ($N_{between}=123$).

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersönlichkeit und Unterrichtsqualität?

Im Fach Mathematik lassen sich auf Individualebene zunächst ähnliche Ergebnisse beobachten wie im Fach Deutsch: Die Dimensionen *Gewissenhaftigkeit*, *Offenheit* und *Verträglichkeit* sind auch hier die Dimensionen, die die meisten signifikanten sowie ausgeprägtesten Einflüsse auf verschiedene Prozessmerkmale, v.a. auf den *Alltagsbezug*, das *Interaktionstempo* und die *Motivierung*, zeigen. Ihnen folgt die Dimension *Resilienz* mit drei zwar signifikanten, jedoch sehr schwach ausgeprägten Effekten (0,07-0,09), die sich nach der Modifikation (Modell 1b) jedoch nicht mehr als signifikant nachweisen lassen. Die individuelle Wahrnehmung der *Extraversion* der Lehrkraft durch ihre Schülerinnen und Schüler scheint keinerlei Wirkung auf deren individuelle Einschätzung der Unterrichtsqualität entlang der Prozessmerkmale zu haben.

Die erklärten Varianzen der Modelle 1a und 1b unterscheiden sich auf Individualebene kaum. Zudem weisen auch jene des Modells mit einem übergeordneten Prädiktor (Modell 3) vergleichbare Werte auf – dementsprechend kann hier von einem Multikollinearitätsbias Abstand genommen werden.

Auf Klassen- resp. Lehrkräfteebene zeigen sich in Bezug auf die Dimension *Gewissenhaftigkeit* kaum Unterschiede zur Individualebene, die Effekte steigen lediglich deutlich an. Die *Extraversion* weist im saturierten Modell (1a) – ganz entgegen der Befunde im Fach Deutsch sowie jener der Individualebene – neben einem starken Effekt auf die Wahrnehmung der *Störungsprävention* auch moderate Effekte auf die *Erklärkompetenz* und die *Motivierungsqualität* sowie einen moderaten negativen Effekt auf die *Methodenvariation* auf. Allerdings verschwinden diese Effekte gänzlich nach der Modifizierung (Modell 1b), lediglich ein schwach ausgeprägter Effekt (0,21) auf die *Störungsprävention* bleibt bestehen. In Bezug auf die Dimension *Offenheit* zeichnen sich andere Effekte ab: Es lassen sich nur noch Effekte auf drei Merkmale (gegenüber fünf auf Individual-ebene) ausmachen, wobei jene auf den *Alltagsbezug* und die *Methodenvariation* in Modell 1a stark und jener auf die *Motivierungsqualität* moderat ausfallen. In Modell 1b zeigen sich alle drei Effekte moderat ausgeprägt. Auf Klassenebene verschiebt sich zudem die *Resilienz* hin zu einem signifikanten und moderat ausgeprägten Effekt auf die *Methodenvariation* sowie einem starken Effekt auf die *Disziplin* (die keine Signifikanz auf Individualebene aufweist). Der Unterricht von allgemein als psychisch stabil geltenden Mathematiklehrkräften scheint demnach methodisch als variantenreicher und disziplinierter wahrgenommen zu werden. Die Effekte auf die *Motivierungsqualität* und die *Erklärkompetenz* sind

Forschungsfrage 2: Welcher Zusammenhang besteht zwischen den Konstrukten
Lehrerpersönlichkeit und Unterrichtsqualität?

auf beiden Ebenen zwar signifikant aber schwach ausgeprägt, jener auf die Motivierung steigt durch die Modifizierung zu einer knapp moderaten Ausprägung an. Die Wirkungsweise der durch die Schülerinnen und Schüler insgesamt wahrgenommenen *Verträglichkeit* der Lehrkraft verhält sich im Fach Mathematik ähnlich wie im Fach Deutsch: Die Signifikanz der Wirkkraft bleibt zwar auf Klassenebene bestehen, dreht sich aber – außer hinsichtlich des *angemessenen Interaktionstempos* – ins Negative.

Die Werte der erklärten Varianz fallen im Fach Mathematik in Modell 1b zwar ab, außer bei der *Methodenvariation* allerdings nicht derart deutlich wie im Fach Deutsch. Hinsichtlich der *Erklärkompetenz* steigt die Varianzaufklärung (R^2) sogar leicht an. Im Vergleich mit der erklärten Varianz aus Modell 3 zeichnen sich äquivalente Tendenzen wie im Fach Deutsch ab: Bis auf die *Störungsprävention* und etwas abgeschwächer die *Disziplin* und das *Interaktionstempo* unterstützen die Anteile der erklärten Varianz aus Modell 3 v.a. das saturierte Modell (1a). Die Werte dieser Merkmale sind demnach mit Bedacht und nur mit plausibler theoretischer Begründung zu interpretieren, da sie möglicherweise durch die Multikollinearität der Prädiktoren resp. der Persönlichkeitsdimensionen beeinträchtigt sind.

IV. Diskussion

17 Interpretation der ermittelten Forschungsbefunde

Trotz bislang mangelhafter entsprechender Erhebungen (und damit einhergehenden unbefriedigenden Ergebnissen)¹⁴⁶, sind in der Schul- und Unterrichtsforschung deutliche Äußerungen zur Bedeutsamkeit der Lehrerpersönlichkeit für die Unterrichtsqualität und damit einhergehend für das Lernen von Schülerinnen und Schülern weit verbreitet (siehe bspw. Helmke 2009, 2014, Hattie 2013, Kunter 2013, Henoch et al. 2015 u.v.m.). Angesichts der bisherigen Forschungslage erscheinen derartige Konstatierungen fragwürdig. Dies noch erheblicher, als das Konstrukt der Persönlichkeit innerhalb der pädagogisch-psychologischen Empirie einem doppelten Forschungsstand ausgesetzt ist, der bislang eher verfestigt denn aufgelöst wurde. So bleiben Untersuchungen zur Lehrerpersönlichkeit bis auf wenige Ausnahmen der Perspektive der Selbstberichte verhaftet, obgleich diesem Vorgehen sowohl theoretische Erkenntnisse als auch Befunde der Differential- resp. Persönlichkeitspsychologie zur perspektivenspezifischen Wahrnehmung von Persönlichkeit sowie deren fluider Charakter entgegenstehen. In Bezug auf die Unterrichtsqualität folgt man hingegen nach einschlägigen Befunden von Clausen (2002) und Gruehn (2000) schon länger dem perspektivenspezifischen Ansatz, was sich in der immer stärkeren Etablierung von Schülerfeedbacks äußert. Vor allem angesichts des vermehrten Einsatzes solcher Schülerfeedbacks gewinnt die Präzisierung von Evaluationen durch Schülerurteile an Bedeutung. Allerdings findet sich kein systematisch befriedigender Ansatz, der die tatsächliche Bedeutung der Lehrerpersönlichkeit für die Schülerinnen und Schüler und deren Unterrichtswahrnehmung untersucht. An diesem Punkt setzt die vorliegende Arbeit an.

Um Einblicke in die Lehrerpersönlichkeit und deren Bedeutung für die Unterrichtsqualität zu gewinnen, wurden über 2000 Schülerinnen und Schüler der Jahrgangsstufe 11 sowie über 100 ihrer Lehrerinnen und Lehrer an gymnasialen Oberstufen (von Berufs- und Gesamtschulen) in NRW per Paper-Pencil-Verfah-

¹⁴⁶ Ein Umstand, an dem sich auch seit identischen Feststellungen von Wachtel (1998) sowie von Seidel und Shavelson (2007) kaum etwas geändert hat.

ren (standardisierte Fragebögen) befragt. Die Einschätzungen von Lehrerpersönlichkeiten wurden sowohl durch die Schülerinnen und Schüler (Fremdeinschätzungen) als auch deren Lehrerinnen und Lehrer (Selbsteinschätzungen) in den Fächern Deutsch und Mathematik vorgenommen und einander zugeordnet. Sie erfolgten zudem entlang des differenzierten und etablierten Konstrukts der Big-Five der Persönlichkeit – also in den Dimensionen *Gewissenhaftigkeit*, *Extraversion*, *Offenheit*, *Resilienz* und *Verträglichkeit*. Im Weiteren wurden die Schülerevaluationen der Unterrichtsqualität erfasst, ebenfalls in ausdifferenzierter Form entlang diverser in der Unterrichtsforschung gängiger und gut geprüfter Prozessmerkmale des Unterrichts – so die *Methodenvariation*, die *Motivierung*, das *angemessene Interaktionstempo*, der *Alltagsbezug*, die *Erklärkompetenz*, die *Störungsprävention* und die *Disziplin*.

17.1 Interpretation des Vorbefunds

Insgesamt zeigt sich, dass das BigFive-Instrument zur Einschätzung der Lehrerpersönlichkeit durch deren Schülerinnen und Schüler im Ganzen gute bis akzeptable Fit-Werte aufweist, auch über die Individualebene hinaus, was angesichts recht hoher Intraklassenkorrelationen zu vermuten war. Die bisweilen eher schwach ausgeprägte Reliabilität mag der dem BigFive-Konstrukt immanenten Heterogenität sowie der Komplexität des Gesamtmodells geschuldet sein und ist weniger geeignet, Rückschlüsse auf mögliche Schwächen des Grundkonstrukts der BigFive zu ziehen. In Anbetracht und unter Abwägung der verschiedenen zu berücksichtigenden Besonderheiten des Konstrukts scheint es dennoch gerechtfertigt, dieses Messmodell innerhalb der weiteren Berechnungen sowohl auf Individual- als auch auf Klassenebene zu verwenden.¹⁴⁷

¹⁴⁷ Anhang B belegt zudem, dass Modelle, die lediglich die einzelnen Skalen der Dimensionen und nicht das Gesamtkonstrukt der BigFive einbeziehen, weitaus bessere und gute Fit-Werte erbringen. Demnach ist in Interpretation der Fit-Werte des Gesamtkonstrukts davon auszugehen, dass diese sich vor allem durch die Modellkomplexität denn durch eine tatsächlich schlechte Passung der Daten mit dem Modell erklären.

Wie akkurat eine Fremdeinschätzung der Persönlichkeit erfolgt, hängt wesentlich von den bereits in Kap. 3.4.1 berichteten Aspekten der Beurteilungskompetenz¹⁴⁸, der generellen Beurteilbarkeit¹⁴⁹, den zu beurteilenden Merkmalen¹⁵⁰ sowie Beurteilungsfehlern¹⁵¹ ab. Die vorgefundenen Model-Fit-Werte lassen den Rückschluss zu, dass Schülerinnen und Schüler durchaus kompetente Beurteiler in der Fremdeinschätzung der Lehrerpersönlichkeit darstellen, zumal sowohl Bekanntheitsgrad mit der Lehrkraft am Ende des Schuljahres (vgl. Kap. 6.2) als auch der Kontext der Beziehung – im schulischen Rahmen – durchaus als stabil verortet werden können. Verzerrungen durch Beurteilungsfehler wird insofern entgegen gewirkt, als die Mehrebenenstruktur berücksichtigt und individuelle Standardfehler korrigiert werden, doch auch auf Individualebene zeigen sich die Schülerantworten reliabel und valide trotz Beobachterbias (vgl. Kap. 4.4.1).

17.2 Interpretation der korrelativen Befunde

Forschungsbefunde berichten vornehmlich von insgesamt eher schwachen bis wenigen mittleren Korrelationen zwischen Selbst- und Fremdeinschätzungen der BigFive-Dimensionen, die je nach Dimension sowie je nach Bekanntheitsgrad der Personen variieren können (vgl. dazu in Kap. 3.4.1 und 2.2.3).

Die hier gefundenen Ergebnisse zeigen ebenenübergreifend Korrelationen in Höhe zwischen 0,03 und 0,41, womit sie in Einklang bspw. mit der Untersuchung von Watson, Hubbert und Wiese (2000) stehen, die ebenfalls von weitgehend schwachen Korrelationen berichten (vgl. Kap. 3.4.1). Darüber hinaus bestätigen die vorliegenden Ergebnisse den Befund, dass die schwachen Korrelationen die ausgeprägten deutlich überwiegen: so zeigen sich auf der Lehrkräfteebene (*between*) nur die Korrelationen zwischen den Schüler- und Lehrerperspektiven in der Dimension *Resilienz* im Fach Deutsch sowie in der Dimension *Extraversion* im Fach Mathematik signifikant und moderat ausgeprägt. Im Fach

¹⁴⁸ Die Beurteilungskompetenz beinhaltet das (Vor-)Wissen, die individuellen Fähigkeit sowie die Motivation zur Einschätzung.

¹⁴⁹ Die generelle Beurteilbarkeit bezieht sich auf den Bekanntheitsgrad mit der zu beurteilenden Person und deren Verhaltenskonsistenz.

¹⁵⁰ Hierbei steht v.a. der Kontext der Beziehung im Vordergrund.

¹⁵¹ Diese gründen v.a. auf impliziten Persönlichkeitstheorien, die durch entsprechende, angemessene Itemkonstruktionen nur teilweise behebbar sind.

Mathematik verfehlt die Dimension *Offenheit* trotz moderater Ausprägung die Schwelle des Signifikanzniveaus knapp.

Beim Vergleich von Selbst- und Fremdbberichten wird in gängigen Forschungen weitgehend die Annahme vertreten, dass starke Zusammenhänge für eine erfolgreiche Selbstpräsentation stehen¹⁵² (Baumeister 1992, DePaulo 1992, Schaffhuser/Allemand/Martin 2014, Rauthmann 2017). Die Ergebnisse werden hier jedoch vor einem anderen, empirisch gewonnenen Hintergrund diskutiert: Da beide Perspektiven gemäß der Überprüfungen aus Kap. 11.3-11.4 als ausreichend reliabel und valide zu bezeichnen sind (mit Ausnahme der Dimension *Verträglichkeit* in den Lehrerselbsteinschätzungen), lassen sich die insgesamt schwach ausgeprägten Zusammenhangsmaße zwischen den Perspektiven zum einen dahingehend interpretieren, dass die Beziehungen zwischen Lehrkräften und ihren Schülerinnen und Schülern keine vertiefte Bekanntschaft darstellen, da mit zunehmender Bekanntschaft die Übereinstimmungen der Wahrnehmung von Persönlichkeitseigenschaften steigen (vgl. Asendorpf/Neyer 2012, Back/Vazire 2012, siehe dazu Kap. 3.4.1). Zum anderen kann angenommen werden, dass trotz identischer Itemformulierungen Unterschiedliches gemessen wird, wonach Mayr und Neuweg (2006) zufolge, aufgrund der breiteren Erfahrungsgrundlage der Lehrkräfte, deren Einschätzungen valider seien als die Schülereinschätzungen. Angesichts der theoretischen Überlegungen zum Konstrukt Persönlichkeit (siehe Kap. 3) in Zusammenhang mit der Perspektive (Kap. 1.1) ist diese Interpretation jedoch kaum haltbar. Es wird hier eher davon ausgegangen, dass nach Clausen (2002) unterschiedliche Perspektiven dieselben Konstrukte unterschiedlich wahrnehmen und dennoch jeweils als valide gelten können (vgl. auch Hofstee 1994, Vazire 2010). Demnach nehmen Schülerinnen und Schüler ihre Lehrerinnen und Lehrer tatsächlich deutlich anders wahr als diese sich selbst. Da Korrelationsmatrizen zwischen fremd- und selbstwahrgenommenen Persönlichkeitsmerkmalen kaum Zusammenhänge zeigen, ließe sich außerdem schlussfolgern, dass sich Intention und Ausdruck der Lehrkräfte nicht entsprechen (vgl. De Paulo 1992), wobei zu beachten ist, dass Fremdbberichte in mancher Hinsicht – d.h. je nach Merkmal – das tatsächlich zur Schau gestellte Verhalten zuverlässiger prognostizieren als

¹⁵² Bei Untersuchungen zu Übereinstimmungen von Fremd- und Selbsteinschätzungen wird davon ausgegangen, dass neben dem Bekanntheitsgrad grundsätzlich auch die Signalgebung eines Merkmals, deren Sichtbarkeit und Wahrnehmung sowie ihre adäquate Interpretation relevant für die Höhe von Korrelationen sind (Rauthmann 2017).

Selbstberichte (Back/Vazire 2012). Neben der Selbstwahrnehmung ist daher die Fremdwahrnehmung bedeutsam, um spezielle Informationen darüber zu erhalten, wie die individuelle Persönlichkeit in spezifischen Umgebungen ausgedrückt, wahrgenommen und eingeschätzt wird (vgl. Kap. 3.4.2).

Des Weiteren erklären sich die schwachen Korrelationen möglicherweise auch dadurch, dass in der Erhebung der Selbsteinschätzung der Lehrerpersönlichkeit der Kontextbezug nicht ausdrücklich berücksichtigt wurde. Angesichts dessen, dass der Kontext maßgeblich zu einer Steigerung von Zusammenhängen verschiedener Wahrnehmungen von Persönlichkeit führt und die Lehrerinnen und Lehrer der vorliegenden Untersuchung kontextfrei befragt wurden, wäre dies (auch angesichts der schwächeren Reliabilitäten) als Grund für die sehr schwache Lehrer-Schüler-Korrelationen interpretierbar. In einer nächsten Untersuchung wird empfohlen, den schulischen Kontext in die Erhebung einzubeziehen und statt „ich bin ein Mensch, der...“ nach „als Lehrer/in, bin ich ein Mensch, der...“ die Persönlichkeitsbefragung einzuleiten. Im Gegensatz dazu ist der Kontext der Schülerbefragung mit der Einleitung „unser Lehrer/unsere Lehrerin...“ deutlich umgrenzt. Abschließend muss festgehalten werden, dass die Kongruenz der Einschätzungen möglicherweise abhängig von dem zu beurteilenden Merkmal ist, was sich je nach Dimension auf die Höhe der Korrelationen zwischen den Perspektiven auswirken kann. Manche Aspekte sind – je nach Beobachtbarkeit und Sichtbarkeit des Merkmals – für andere, manche für den/diejenige/n selbst besser einsehbar. Je sichtbarer ein Merkmal ist, desto kongruenter zeichnen sich die Perspektiven in der Regel ab (vgl. Kap. 3.4.1, vgl. Back/Vazire 2012, Watson/Hubbert/Wiese 2000). Entsprechend wird die Dimension *Neurotizismus* Vazire (2010) zufolge in der Selbsteinschätzung am akkuratsten erfasst, obwohl in Bezug auf die Dimension *Neurotizismus* gilt, dass Selbsteinschätzungen dieser Dimension eher Teil des eigenen *Neurotizismus* sind, als dass sie tatsächlich Ausdruck in spezifischen sozialen Interaktionen finden (vgl. Asendorpf/Neyer 2012). *Extraversion* und *Offenheit für neue Erfahrungen* können hingegen als gut sichtbare Merkmale sowohl selbst als auch von nahen und entfernten Bekannten adäquat eingeschätzt werden¹⁵³ (Vazire 2010, vgl. auch Mayr 2019). Über verschiedenste Beurteilergruppen hinweg zeigen insgesamt und ohne Messfehlerbereinigung

¹⁵³ Das SOKA-Modell von Vazire (2010) soll bspw. darüber aufklären, welche Persönlichkeitsaspekte besser selbst und welche besser von anderen eingeschätzt werden.

die Dimensionen *Extraversion* und *Neurotizismus* die höchsten Beurteilerübereinstimmungen (vgl. Rauthmann 2017).

In der vorliegenden Untersuchung ist im Fach Deutsch die Dimension *Neurotizismus* resp. *Resilienz* diejenige, die mit Abstand die höchste (0,41) sowie die einzige signifikante Korrelation zwischen Lehrer- und Schülerperspektiven aufweist. Erklären ließe sich dies dadurch, dass durch die Besonderheiten der Arbeits- und Organisationsstrukturen sowie der eher auf den Ausdruck, die Wahrnehmung und die Interpretation von Meinungen und Gefühlen zielende Kommunikation (vgl. Labudde 2009) im Fach Deutsch entweder die Schülerinnen und Schüler auf neurotizistische Züge ihrer Deutschlehrkräfte besonders sensibilisiert sind oder die entsprechenden Merkmale im Deutschunterricht der Oberstufe bei den Lehrkräften besonders deutlich zutage treten. Plausibel erscheint zudem, dass die Schülerinnen und Schüler die Itemformulierungen in dieser Dimension am ehesten in identischer Weise wie ihre Lehrkräfte interpretieren.

Im Fach Mathematik hingegen zeigen sich lediglich in der Dimension *Extraversion* signifikante Korrelationen zwischen den Selbsteinschätzungen der Lehrkräfte und den Fremdeinschätzungen ihrer Schülerinnen und Schüler. Diese bleiben auch bei Kontrolle für das Lehrkraftgeschlecht bestehen. Dass die Schülerinnen und Schüler die Itemformulierungen in dieser Dimension in weitgehend identischer Weise interpretieren wie es ihre Lehrkräfte tun, könnte auch hier einen Erklärungsansatz bieten. Dasselbe gilt für die Möglichkeit, dass die entsprechenden Merkmale der Dimension *Extraversion* im Mathematikunterricht der Oberstufe bei den Lehrkräften besonders deutlich zutage treten und von den Schülerinnen und Schülern adäquat interpretiert werden.

Denkbar ist auch, dass Lehrer-Schüler-Beziehungen derart von Beziehungen im Rahmen von Freundschaft oder Bekanntschaft abweichen, dass sich dadurch gänzlich eigenständige Korrelationsstrukturen ergeben. Wurde diesbezüglich in früheren Studien die unterrichtsmethodische Grundeinstellung der Lehrkraft als Kovariate in Korrelationen einbezogen, sanken bei Lehrkräften mit einer konstruktivistischen Orientierung die Korrelationen zwischen Lehrerselbst- und Schülerfremdeinschätzungen in den Dimensionen *Extraversion* und *Offenheit für neue Erfahrungen* auf 0,40 (*Extraversion*) und 0,32 (*Offenheit für neue Erfahrungen*) (vgl. hierzu Mayr/Neuweg 2006, Mayr 2019). Dies könnte möglicherweise für stark am schulischen bzw. unterrichtlichen Kontext orientierte Schülereinschätzungen sprechen.

An diesem Punkt ist außerdem zu beachten, dass laut Forschungsbefunden etwa 40% der Lehrerinnen und Lehrer aussagen, dass sie den Schülerinnen und Schülern gegenüber ihre Gefühle und Gedanken nicht offen zeigen (Randoll 1995). Zwar zeigt demnach der überwiegende Teil zumindest meistens „was sie denken und fühlen“ (Randoll 1995, S. 21), allerdings wäre denkbar, dass dies zu Verzerrungen bei den Schülerwahrnehmungen führt und die Korrelationen vor diesem Hintergrund zu interpretieren sind.

17.3 Interpretation der regressionsanalytischen Befunde

Drei Befunde treten bei den Regressionsanalysen deutlich hervor: Erstens fallen die teils enorm hohen Anteile ($R^2 \geq 0,6$) erklärter Varianz der Prozessmerkmale durch schülerperzipierte Persönlichkeitseigenschaften der Lehrkräfte auf, v.a. auf Klassen- resp. Lehrkräfteebene. Diese zeigen sich im saturierten, im spezifizierten sowie in jenem Modell mit einem übergeordneten Persönlichkeitsfaktor 2. Ordnung. Zweitens ergibt sich auf Klassenebene ein negativ ausgeprägter Effekt der *Verträglichkeit* (der sich demgegenüber auf Individualebene schwach positiv abzeichnet). Drittens bestehen in den Effekten zwar durchaus Parallelen zwischen den beiden Fächern Deutsch und Mathematik, jedoch lassen sich auch deutliche Fachunterschiede in der Wirkungsweise der Lehrer-BigFive (aus Schülersicht) beobachten.

In Bezug auf den ersten Punkt kann angenommen werden, dass die Prozessmerkmale *Erklärkompetenz* und *Motivierung* eher persönliche Dimensionen der Lehrkraft messen als Unterrichtsmerkmale.¹⁵⁴ Dies ist zudem inhaltlich plausibel begründbar, da beide Merkmale bei ihrer Operationalisierung sprachlich deutlich an die Person der Lehrkraft gekoppelt sind. Um bspw. zu verstehen, was ein Lehrer oder eine Lehrerin zu erklären versucht bzw. um einen kognitiven Inhalt klar und strukturiert an die Schülerinnen und Schüler zu vermitteln, bedarf es des persönlichen Kontakts zwischen den Schülerinnen und Schülern und ihren Lehrerinnen und Lehrern. Ebenso basiert *Motivierung* vornehmlich auf zwischenmenschlichen, persönlichen Aspekten. Dabei erfordert eine hohe Qualität der *Motivierung* bspw. einen positiven Bezug zur motivierenden Person, sofern keine intrinsische Motivation, geprägt durch *individuelles Interesse*, besteht (vgl.

¹⁵⁴ Hier sei auf Befunde von Gruehn (2000) verwiesen, die in ihrer Untersuchung ähnliche Beobachtungen bei diesen beiden Skalen machte.

Deci/Ryan 2000). Allerdings zeigen sich nach der Modellspezifikation unterschiedliche Ausprägungen der Varianzaufklärung je nach Fach: Während in Mathematik die Varianz nahezu unverändert in hohem Maße ($R^2=0,81$ bzw. $R^2=0,73$) durch Dimensionen der Lehrkraftpersönlichkeit erklärt wird, sinkt im Fach Deutsch nach der Modellspezifikation der Anteil erklärter Varianz hinsichtlich des Merkmals *Motivierung* deutlich ab (von $R^2=0,80$ auf $R^2=0,31$). Daraus ließe sich schließen, dass die *Motivierungsqualität* in Mathematik stärker an die Person der Lehrkraft gebunden ist.

Beiden Fächern gleich ist demgegenüber – bezogen auf den zweiten Punkt – bspw. die Tendenz der Effekte der schülerperzipierten *Verträglichkeit* der Lehrkräfte auf die Prozessmerkmale *Disziplin* und *Störungsprävention*: Wirkt sich die wahrgenommene *Verträglichkeit* auf individueller Ebene sehr schwach (wenn auch signifikant) auf die *Disziplin* und nicht signifikant auf die *Störungsprävention* aus, gewinnt sie auf Klassenebene in beiden Fächern und hinsichtlich beider Merkmale deutlich an Effektstärke; auf die *Störungsprävention* noch ausgeprägter als auf die *Disziplin*. Dieser Effekt auf Klassenebene ist darüber hinaus negativ ausgeprägt. Daraus ließe sich ableiten, dass Lehrkräfte, die innerhalb der Klasse als nicht gut verträglich gelten, positiv hinsichtlich ihres disziplinarischen und störungspräventiven Umgangs mit der Klasse gewertet werden. Möglicherweise führt der Ruf einer Lehrkraft, eher unfreundlich und gröber im Umgang sowie nachtragend zu sein,¹⁵⁵ zu einer höheren Regelklarheit, Störungsfreiheit und vor allem Zeitnutzung im Unterricht¹⁵⁶. Denkbar wäre zudem, dass sich in diesen Effekten der Respekt (oder auch die Furcht) der Schülerinnen und Schüler vor einem unerwünschten Tadel vor der gesamten Klasse bzw. im Kontext der Klasse ausdrückt. Störungen des Unterrichts und Tadel betreffen plausiblerweise eher die gesamte Klasse bzw. erfolgen innerhalb ihres Kontexts (in Form von Interaktionen), denn auf individueller Ebene mit dem Lehrer oder der Lehrerin. Entsprechend sind auf Individualebene diese Zusammenhänge unwesentlich bzw. nicht von Bedeutung. Darüber hinaus zeigt sich die *Verträglichkeit* wiederum für die Wahrnehmung einer *Angemessenheit des Interaktionstempos* auf In-

¹⁵⁵ Diese Facetten der Dimension *Verträglichkeit* werden im BFI-S vornehmlich erhoben.

¹⁵⁶ Auf diese Aspekte ist die Erfassung der beiden Merkmale *Disziplin* und *Störungsprävention* v.a. ausgerichtet.

dividual- wie auch auf Klassenebene als bedeutsam. Ein und dieselbe persönliche Eigenschaft wirkt demnach offenbar nicht auf alle Prozessmerkmale gleichermaßen.

Angesichts der Tatsache, dass Untersuchungen zur Prädiktionskraft persönlicher Merkmale generell und innerhalb der Unterrichtsforschung im Besonderen fast ausschließlich auf Selbsteinschätzungen der Lehrkräfte beruhen, bieten nur vereinzelte Studien Erfahrungswerte, die einer Interpretation der hier vorgefundenen Ergebnisse zur Orientierung dienen können. Denn die meisten dieser Studien basieren entweder auf kleinen Stichproben innerhalb amerikanischer Universitäten, die zudem kaum mit dem schulischen Unterricht hierzulande vergleichbar sind, oder erfolgen in derart unsystematischer Weise, dass keine vergleichbaren Konstrukte und damit Erfahrungswerte abgeleitet werden können. Eine der wenigen Ausnahmen bildet die Metaanalyse Feldmanns (1986). Sein Befund, dass Fremdeinschätzungen der Lehrerpersönlichkeit mit Unterrichtseinschätzungen aus Schülerperspektive durchaus korrelieren (ausführlich in Kap. 2.2), wird in der hier durchgeführten Untersuchung weitgehend bestätigt. Dem Fazit Feldmanns, dass es für die Bedeutung des Zusammenhangs von wahrgenommenen Merkmalen der Lehrerpersönlichkeit mit der Unterrichtsqualität angebracht scheint, verschiedene Perspektiven auf die Lehrerpersönlichkeit getrennt voneinander zu analysieren, wird in dieser Arbeit ausdrücklich gefolgt. Zumal Feldmann den Wahrnehmungen der Schülerinnen und Schüler zwar Validität zugesteht, allerdings unter der – auch in dieser Arbeit gegebenen – Prämisse, dass es sich dabei nicht um die Beschreibung einer *wahren* Persönlichkeit handelt, sondern lediglich um eine spezifische und kontextuelle Perspektive. Allerdings lässt sich auch in der vorliegenden Untersuchung nicht ausschließen, dass Persönlichkeitseinschätzungen ebenso wie Unterrichtseinschätzungen durch die Schülerinnen und Schüler denselben Biasvariablen unterliegen und sich dadurch die Höhe der Korrelationen erklären lässt (vgl. Feldmann 1986 und Kap. 4.4.1).

Weitere Untersuchungen weisen nach, dass die mittels Peer-Ratings¹⁵⁷ erfasste Lehrerpersönlichkeit einen signifikanten Prädiktor für effektiven Unterricht bzw. für positive Schülerratings bzgl. der Lehreffektivität darstellen kann (Schmidt et al. 1991, Murray et al. 1990, Magno/Sembrano 2008). Renaud und

¹⁵⁷ Überwiegend ist hier gegenseitiges Peer-Rating gemeint: Dieses gilt als intern konsistenter und stärker mit Lehreffektivität korreliert als Selbst- oder Schülereinschätzung der Lehrerpersönlichkeit (Feldmann 1986).

Murray (1996) zeigten diesbezüglich in einer Untersuchung, die auf Peer-Ratings von 33 Psychologieprofessoren einer Fakultät beruht, dass die Lehreffizienz signifikant mit einzelnen Persönlichkeitseigenschaften korreliert, nicht aber mit *Extraversion* oder *Neurotizismus*. Eine weitere Studie (von Clayson/Sheffet 2006, Details siehe Kap. 2.2.2) stellt heraus, dass Persönlichkeitsdimensionen – wiederum außer *Extraversion* – signifikante Regressionskoeffizienten in Bezug auf einen übergeordneten Faktor *Evaluation* aufweisen, sie also starke Prädiktoren sind. Ebenso zeigt die Studie, dass die *Unterrichtsevaluation* mit Veränderungen der Wahrnehmung der Lehrerpersönlichkeit variiert, dies allerdings v.a. hinsichtlich der Wahrnehmung von *Extraversion*. Korrelationen zwischen dem übergeordneten Persönlichkeitsfaktor und der erhobenen Unterrichtsevaluation nehmen zudem über vier Erhebungszeitpunkte immer stärker zu, wobei sie durchweg signifikant sind. Im Fach Deutsch können die Befunde hinsichtlich der BigFive-Dimension *Extraversion* bestätigt werden: Hier lassen sich keine Effekte auf die Unterrichtsqualität, gemessen anhand der verschiedenen Prozessmerkmale, verzeichnen. Die *Resilienz* hingegen zeigt auf Klassenebene im Fach Deutsch deutliche Effekte auf fünf (von sieben) Prozessmerkmale. Allerdings ist dieser Befund keineswegs fachübergreifend: Im Fach Mathematik zeigt die Dimension *Resilienz* drei moderate und einen knapp starken Effekt auf die Prozessmerkmale des Unterrichts. *Extraversion* stellt im Fach Mathematik jedoch ebenso wie die *Resilienz* im Fach Deutsch deutliche Prädiktoren (0,26-0,67) für vier von sieben Prozessmerkmalen des Unterrichts. Dies gilt jedoch nur für das saturierte Modell (1a), während die Effekte im spezifizierten Modell 1b nicht signifikant bzw. in Bezug auf die *Störungsprävention* schwach ausgeprägt sind. Angesichts dieser Befunde ließe sich schlussfolgern, dass die Ergebnisse der oben berichteten Studien u.U. durch eine fachübergreifende Betrachtung zustande kommen.

Vor dem Hintergrund der Studie von Klusmann et al. (2006, siehe hierzu den entsprechenden Abschnitt in Kap. 2.2), die im Fach Mathematik anhand von selbsteingeschätzten Lehrerbelastungstypen nach AVEM deutliche Effekte des positiven und als resilient zu bezeichnenden Gesundheitstyps auf vier von sechs Prozessmerkmalen der von den Schülerinnen und Schüler wahrgenommenen Unterrichtsqualität ausmachen konnten, lassen sich die hier beobachteten Effekte der *Resilienz* interpretieren. Zum einen dahingehend, dass die Wirkung der *Resilienz* der Lehrkräfte auf die schülerperzipierte Unterrichtsqualität trotz differenter Erhebungsformen Ähnlichkeiten aufweist: Sowohl die im Lehrerselbstbericht

erfassten AVEM-Typen als auch die in der vorliegenden Untersuchung durch die Schülereinschätzungen erfasste Lehrerpersönlichkeit entlang der BigFive zeigen keine Effekte der *Resilienz* der Lehrkräfte auf solche unterrichtlichen Aspekte, die sich auf Störungen im Unterricht beziehen, jedoch durchaus auf andere Unterrichtsaspekte (siehe „Störungen im Unterricht“ bei Klusmann et al. (2006) resp. *Störungsprävention* in der vorliegenden Arbeit). Zum anderen lässt sich möglicherweise die BigFive-Dimension *Resilienz* vergleichbar oder ggf. noch differenzierter anhand der AVEM-Typen erfassen. Da aus Schülersicht die Unterrichtsqualität von ein und derselben Lehrkraft fachübergreifend variieren kann und zudem gewisse persönliche Merkmale offenbar in einem Fach bedeutsam für die Unterrichtsqualität und im anderen bedeutungslos sein können (vgl. auch Murray et al. 1990), lassen sich die Ergebnisse von Klusmann et al. (2006) jedoch nicht auf andere Fächer übertragen. Dagegen deuten die Befunde der vorliegenden Untersuchung hinsichtlich der schülerperzipierten *Resilienz* der Lehrkräfte vergleichbare Tendenzen in den Fächern Deutsch und Mathematik an.

Bemerkenswert ist an dieser Stelle, dass in beiden Fächern auf beiden Ebenen (individuell und klassenbezogen) vor allem die schülerperzipierten Dimensionen *Gewissenhaftigkeit* und *Verträglichkeit* als dominant herausstechen, zudem (in Deutsch leicht verstärkt) die *Resilienz* und v.a. in Mathematik die *Offenheit*. Jene Dimensionen, die vermehrt für die Feststellung der Berufseignung herangezogen werden (bspw. in Assessment Centers, siehe hierzu Kap. 2.2.4), sind hingegen die selbsteingeschätzte *Gewissenhaftigkeit*, *Resilienz* und *Extraversion* (vgl. Mayr 2003). Zwar vermögen diese Dimensionen sicherlich etwas über den Studienerfolg aussagen; gemäß den Befunden der vorliegenden Untersuchung jedoch kaum etwas über den beruflichen Erfolg von Lehrerinnen und Lehrern, sofern dieser im Sinne hochwertiger Unterrichtsprozesse aus Schülersicht verstanden wird.

Es ist außerdem festzuhalten, dass die Deutlichkeit der Effekte (auf Klassenebene) sowie die hohe Varianzaufklärung möglicherweise auf den systematischen Einsatz etablierter Konstrukte und gut validierter Instrumente zurückzuführen ist, auf dem diese Arbeit basiert und nicht zwangsläufig als Hinweis auf Multikollinearität anzusehen ist. Zumal auch Murray et al. (1990) und Feldmann

(1986) von einer hohen Varianzaufklärung von Schülereinschätzungen des Unterrichts durch Persönlichkeitsmerkmale ausgehen:

„Personality traits taken collectively can account for up to 75% of between-teacher variance in student instructional ratings.” (Murray 1996, S. 324)

Dieser Wert entspricht weitgehend der erklärten Varianz der spezifizierten Modelle (1b). Das saturierte Modell sowie jenes mit einem übergeordneten (Persönlichkeits-)Faktor 2. Ordnung weisen hingegen für mindestens zwei Prozessmerkmale etwas höhere Varianzaufklärungen auf.

18 Zusammenfassende Diskussion

Nachdem zunächst die eingesetzten Instrumente validiert und geprüft wurden, konnte hinsichtlich der Fähigkeit der Schülerinnen und Schüler die Lehrerpersönlichkeit einzuschätzen ein erster Vorbefund herausgestellt werden. Dieser zielt darauf ab, dass Schülerinnen und Schüler durchaus in der Lage sind – vor allem auch weit über rein individuelle Antworttendenzen hinaus – ihre Lehrerinnen und Lehrer zuverlässig mit dem eingesetzten Instrument einzuschätzen. Demnach erscheint es legitim zu fordern, dass in der Forschung nicht ausschließlich Selbsteinschätzungen in Bezug auf die Lehrerpersönlichkeit einbezogen werden, wie es bislang weitestgehend üblich ist, sondern die Schülerperspektiven auf die Lehrerpersönlichkeit gleichwertig einzubinden sind.

Des Weiteren zeigen die Analysen zur Validierung, dass die Schülerinnen und Schüler sowohl die Lehrerpersönlichkeiten als auch die Prozessmerkmale der Unterrichtsqualität deutlich fachspezifisch in Deutsch und Mathematik einschätzen. Forschungen weisen in Zusammenhang mit Einschätzungen des Unterrichts darauf hin, dass zwar durchaus allgemeine Qualitätsmerkmale bestehen, deren Interpretation und Auslegung jedoch je nach Fach variieren. Diese Annahme ließe sich auf die Wahrnehmung der Lehrerpersönlichkeit durch die Schülerinnen und Schüler übertragen, da es wenig plausibel erscheint, dass sich die Lehrerinnen und Lehrer je nach Fach in ihrer Persönlichkeit unterscheiden. Vor allem die Fächer Deutsch und Mathematik unterscheiden sich in ihren Lernzielen und -inhalten sowie ihrer Didaktik, Aufgaben- und Organisationskultur wesentlich (vgl. Gruehn 2000, Labudde 2009).

Die deskriptiven Befunde zeichnen ein insgesamt positives Bild der Lehrermerkmale sowohl im Selbstbericht als auch in den Schülereinschätzungen. So liegen nahezu alle Mittelwerte im oberen Drittel der eher positiv konnotierten Merkmalsdimensionen. Bei den Lehrereinschätzungen zeigt sich lediglich die *Resilienz*, bei den Schülereinschätzungen hingegen die *Offenheit für neue Erfahrungen* insgesamt schwächer ausgeprägt. Im Fach Mathematik ist diese Tendenz etwas deutlicher als im Fach Deutsch. Auch die Mittelwerte der Schülereinschätzungen der Unterrichtsqualität befinden sich gehäuft in der oberen, positiv befundenen Hälfte der Skalierung. Lediglich die *Methodenvariation*, die *Motivierung* und der *Alltagsbezug* werden in Mathematik eher darunter eingeschätzt. Am besten fallen *Interaktionstempo* und *Erklärkompetenz* in beiden Fächern aus, in Deutsch

etwas höher gegenüber Mathematik. Diese Befunde spiegeln insgesamt und weitgehend bestehende Deskriptionen aus der Forschung wider (wobei keine Vergleichsdaten von Schülereinschätzungen von Lehrerpersönlichkeiten entlang der BigFive vorliegen). Dadurch ist anzunehmen, dass die Stichprobe durchaus als repräsentativ angesehen werden darf.

Die Ergebnisse zur ersten Forschungsfrage, inwiefern Lehrerselbst- und Schülerfremdeinschätzungen der Lehrerpersönlichkeit zusammenhängen, wurden über Korrelationen ermittelt. Insgesamt zeigen sich in Hinblick auf die Einschätzungen der Lehrerpersönlichkeit signifikante, wenn auch nur moderate Korrelationen zwischen Schüler- und Lehrerperspektive in den BigFive-Dimensionen *Extraversion* und *Resilienz*, dies zudem fachspezifisch: Genauer betrachtet ergibt sich, dass lediglich die *Resilienz* im Fach Deutsch deutliche Übereinstimmungen zwischen Lehrer- und Schülereinschätzungen anzeigt. Die Signifikanz der Zusammenhänge in der Dimension *Extraversion* hingegen tritt im Fach Mathematik heraus und bleibt auch bei Kontrolle des Lehrkraftgeschlechts bestehen. Unter den Schülerinnen und Schülern ließen sich in dieser Dimension zuvor keine lehrkraftgeschlechtsabhängigen Mittelwertunterschiede feststellen, was jedoch im Fach Mathematik hinsichtlich der *Resilienz* und der *Offenheit* der Fall ist (die korrelationsanalytisch aber nicht wesentlich ins Gewicht fallen). Wie in der Interpretation von Forschungsfrage 1 bereits festgestellt, entsprechen die Befunde bestehenden Erkenntnissen zu Zusammenhängen zwischen Fremd- und Selbsteinschätzungen der Persönlichkeit insofern, als zum einen kaum Zusammenhänge bestehen, und jene in der Dimension *Extraversion* durch die für andere gute Sicht- und Beobachtbarkeit extravertieren Verhaltens plausibel erscheinen (vgl. Kap. 3.4.1). Allerdings fällt die Dimension *Resilienz* mit der einzigen eindeutigen Korrelation im Fach Deutsch aus dem Rahmen der im Forschungsstand berichteten Ergebnisse sowie der theoretisch abgeleiteten Erwartungen heraus. Zumal die Validität der Schülereinschätzungen gerade in dieser Dimension Schwächen aufweist, die den Rückschluss zulassen, dass die Schülerinnen und Schüler genau diese Dimension bei ihren Lehrerinnen und Lehrern möglicherweise nicht präzise zu erfassen vermögen. Dass im Fach Deutsch die Fremdwahrnehmungen der Schülerinnen und Schüler in Bezug auf die *Resilienz* (mit 13% gemeinsamer Varianz) eher mit den Resilienzselbsteinschätzungen ihrer Lehrkräfte übereinstimmen als hinsichtlich anderer BigFive-Dimensionen lässt sich durch die Besonderheiten der Arbeits- und Organisationsstrukturen sowie der eher auf den Ausdruck, die Wahrnehmung und die Interpretation von

Meinungen und Gefühlen zielende Kommunikation (vgl. Labudde 2009) im Fach Deutsch erklären.

Nach dem Befund, dass die Selbsteinschätzung der Lehrerpersönlichkeit korrelativ insgesamt kaum in Zusammenhang mit der schülerperzipierten Lehrerpersönlichkeit oder der schülerperzipierten Unterrichtsqualität steht, was die diesbezügliche Forschungslage weitgehend bestätigt, zeigt sich demgegenüber eine weitaus eindeutiger Struktur in den fachspezifischen Modellen der Schülerperspektiven, die auf Grundlage theoretischer Überlegungen regressionsanalytisch ausgerichtet sind (siehe Forschungsfrage 2). Als erstes wird insgesamt sichtbar, dass nicht nur die Wahrnehmungen der Schülerinnen und Schüler hinsichtlich der Lehrerpersönlichkeit und der Unterrichtsqualität fachspezifisch ausfallen, sondern ebenso die Effekte von Merkmalen der Lehrerpersönlichkeit auf Prozessmerkmale der Unterrichtsqualität zumindest auf Klassenebene je nach Fach variieren. Demzufolge zeichnet sich im Fach Deutsch ein anderes Wirkungsmuster ab als im Fach Mathematik. Vor allem fällt dabei auf, dass auf Klassenebene im Fach Mathematik insgesamt mehr signifikante Effekte von der schülerperzipierten Lehrerpersönlichkeit auf die Unterrichtsqualität ausgehen als im Fach Deutsch (19:13 bzw. nach der Modifikation 16:13). Zudem ist die Varianzaufklärung der wirksamen BigFive-Dimensionen der Lehrerpersönlichkeit im Fach Mathematik tendenziell ausgeprägter als im Fach Deutsch. In Deutsch gehen auf Klassen- resp. Lehrkraftebene signifikante und stark ausgeprägte ($\geq 0,6$) Effekte lediglich von den Dimensionen *Gewissenhaftigkeit*, *Verträglichkeit* und *Resilienz* aus. Die Dimension *Extraversion* hingegen weist keinerlei signifikanten Effekte und die *Offenheit* ausschließlich auf Individualebene moderate Effekte auf. In Mathematik geht von allen Dimensionen – außer auf Schüler- resp. Individualebene der *Extraversion* und im modifizierten Modell der *Resilienz* (die im saturierten Modell bereits sehr schwach ausgeprägt ist) – mindestens ein signifikanter Effekt aus, von der *Verträglichkeit* und (im saturierten Modell) der *Offenheit* sogar zwei. Mit der Modellspezifikation, entlang derer ausschließlich die signifikanten Pfade modelliert werden, lassen sich ebenso wie im Fach Deutsch auch im Fach Mathematik nur noch die Dimensionen *Gewissenhaftigkeit*, *Verträglichkeit* und *Resilienz* mit starken Effekten verzeichnen. Gemeinsam ist beiden Fächern, dass alle Effekte vor allem auf Lehrkraftebene (between) deutlich hervortreten, während auf Individualebene (within) lediglich die *Offenheit* und die *Verträglichkeit* ein bis zwei gleichermaßen moderate ($\geq 0,3$) sowie in Deutsch knapp moderat die *Gewissenhaftigkeit* Effekte zeigen.

In Mathematik, so ließe sich schlussfolgern, bedarf es einer aus Schülersicht etwas facettenreicheren Persönlichkeit, um eine von den Schülerinnen und Schülern positiv empfundene Unterrichtsqualität zu bewirken, als im Fach Deutsch, wo vor allem die wahrgenommene *Gewissenhaftigkeit* und *Verträglichkeit* sowie etwas abgeschwächter die *Resilienz* eine positive Einschätzung der Unterrichtsqualität bedingen. Welche Erklärungen lassen sich für die beiden fachspezifischen Wirkungsmuster heranziehen? Stolz gibt einen entsprechenden Denkanstoß mit der folgenden Aussage:

„Des weiteren ist in Rechnung zu stellen, daß grundsätzlich mögliche Beziehungen zwischen der Lehrerpersönlichkeit und der Wahrnehmung durch die Schüler von situativen Bedingungen durchkreuzt werden können, denn mit Sicherheit modifizieren unterrichtliche Gegebenheiten das Lehrerverhalten.“ (Stolz 1997, S. 136)

Denkbar wäre demnach bspw., dass sich die Fachdifferenzen durch Unterschiede in der Aufgaben- und Organisationsstruktur (Gruehn 2000, S. 209) sowie der Lern-/Bildungsziele und dementsprechend unterschiedlichen didaktischen Vorgehensweisen (vgl. Labudde 2009) erklären lassen. Während sich die Schülerinnen und Schüler im Fach Deutsch und verstärkt in der Oberstufe in die inhaltliche Ebene von Texten vertiefen und dabei kommunikativ den „Austausch zwischen Menschen erfahren; Sachverhalte, Situationen, Meinungen und Gefühle wahrnehmen und interpretieren, ausdrücken und darüber nachdenken“ (Labudde 2009, S. 332), steht im Fach Mathematik „[d]as Vorstellungsvermögen im Umgang mit Zahlen und in der Begegnung mit Objekten und Sachverhalten aus Natur und Mitwelt“ (Labudde 2009, S. 332) im Fokus. Möglicherweise treten bei der unter anderem eher emotional determinierten Beschäftigung gewisse Anteile der wahrgenommenen Lehrerpersönlichkeit in den Hintergrund und andere, möglicherweise auf Empathie beruhende Anteile in den Vordergrund. So ließe sich weiter folgern, dass *Extraversion* und *Offenheit* eher egozentrisch ausgerichtete Merkmale sind, während hohe *Verträglichkeit*, *Gewissenhaftigkeit* und *Resilienz* eher auf soziale Beziehungen abzielen, die vor allem für die Unterrichtsqualität im Fach Deutsch in der Oberstufe relevant sein könnten (vgl. hierzu auch Kap. 4.4.1). Im Fach Mathematik hingegen wird vor allem in der Oberstufe eher auf stark abstrahierter Ebene gearbeitet – Ableitungen, Stochastik und der Umgang mit Formeln tangieren die emotionale Ebene der Schülerinnen und

Schüler womöglich weniger als die Auseinandersetzung mit literarischen Texten, was möglicherweise zu einer gesteigerten und facettenreicheren Bedeutsamkeit der Lehrerpersönlichkeit für die Unterrichtsqualität führt. Um derartige, ans Spekulative grenzende Interpretationen zu untermauern, bräuchte es jedoch weit mehr an spezifisch auf Fachunterschiede ausgerichteter Forschung.

Beiden Fächern gemeinsam ist hingegen, dass sich die Unterrichtsmerkmale *Disziplin* und *Störungsprävention* auf eine negativ wahrgenommene *Verträglichkeit* regredieren lassen. Als eher unverträglich geltende Lehrkräfte bedingen demnach eine hohe *Regelklarheit* und *Zeitnutzung*, während gleichzeitig v.a. die eingeschätzte *Angemessenheit des Interaktionstempos* eines hohen Maßes an *Verträglichkeit* seitens der Lehrkraft bedarf. Dieser Befund ließe sich zum einen dadurch erklären, dass der Gegenpol der *Verträglichkeit* – also sozial eher unverträgliches Verhalten – in Bezug auf Unterricht seitens der Schülerinnen und Schüler möglicherweise als strenges Verhalten der Lehrkraft aufgefasst wird und dadurch zum anderen einem ausgeprägten Maß an Klassenführung zugute kommt. Auf Individualebene hingegen wird ersichtlich, dass die negativen Effekte der *Verträglichkeit* nicht bestehen, im individuellen Kontakt zwischen der Lehrkraft und ihren Schülerinnen und Schülern also ein positives soziales Agieren der Lehrkraft eher förderlich für eine positive Unterrichtswahrnehmung seitens der Schülerinnen und Schüler ist.

19 Fazit

Die Persönlichkeit als relativ stabiles Konstrukt scheint eine Prädisposition für das tatsächliche Verhalten darzustellen (vgl. Guthrie et al. 1998), woraus sich ableiten lässt, dass Lehrerpersönlichkeitseigenschaften die Durchführung des Unterrichts beeinflussen (vgl. Polk 2006, Mayr/Neuweg 2018). So betrachtet ist die Unterrichtsgestaltung insofern von der Persönlichkeit der Lehrkraft beeinflusst, als die Lehrkraft viele Entscheidungen nach ihren Neigungen fällen kann und muss (vgl. auch Rubin/Bruce 1985). Allerdings zeigte sich, dass Schülerwahrnehmungen der Lehreffektivität stärker direkt mit Lehrerpersönlichkeitsmerkmalen als mit dem Lehrerverhalten verknüpft sind (vgl. Hart/Driver 1978, Sherman/Blackburn 1974). Hinzu kommt der Befund, dass Einschätzungen älterer¹⁵⁸ Schülerinnen und Schüler von der Kompetenz ihrer Lehrkraft sowie ihrer Sympathie mit der Lehrkraft faktoriell nicht differieren. Schülerinnen und Schüler scheinen diese Konstrukte also nicht voneinander zu unterscheiden (vgl. Jones 1989), während sie zugleich die Persönlichkeit als wesentlichen Faktor für ihr Lernen ansehen (vgl. Lublin/Barrand 1987). Der Großteil bisheriger empirischer Forschungen zur Bedeutung der Lehrerpersönlichkeit für die Unterrichtsqualität lässt allerdings kaum valide Rückschlüsse zu, gleichzeitig fallen Studien, die auf Umwegen versuchen, die Effekte der Lehrerpersönlichkeit zu erfassen (wie die oben erwähnten), immer wieder auf die direkte Bedeutung der Lehrerpersönlichkeit zurück. Das legt nahe, den Mangel eindeutiger Ergebnisse zum einen mit unsystematischen Untersuchungsdesigns (kleine und/oder selektive Stichproben) und zum anderen Indifferenzen bzgl. der Perspektive der Persönlichkeitserfassung zu erklären (vgl. Kap. 3). Zwar wird die Annahme, dass das Lehrerverhalten während des Unterrichtsprozesses auf der Lehrerpersönlichkeit gründet, immer wieder explizit betont. Im Gegensatz zur Prozessqualität wird die Lehrerpersönlichkeit jedoch generell im Selbstbericht erhoben. Mit den Ergebnissen von Clausen (2002), die Wahrnehmungen vor allem perspektivenspezifisch valide mit wenigen perspektivenübergreifenden Zusammenhängen darstellen, lässt sich schließen, dass die perspektivenübergreifende Erhebungsform in weiten

¹⁵⁸ In der Untersuchung handelt es sich um College-Studierende.

Teilen ursächlich für die Inhomogenität innerhalb der Schulforschung und ihrer Befunde zum Zusammenhang von Lehrermerkmalen und Lehrerverhalten ist. Anders verhält es sich in Bezug auf den Einsatz von Persönlichkeitserfassungen innerhalb von Assessment Centern. Studien als auch Metaanalysen (Tokar/Fischer/Subich 1998, Judge/Heller/Mount 2002) stellen heraus, dass die BigFive-Dimensionen *Extraversion*, *Resilienz* und *Gewissenhaftigkeit* für das berufliche Befinden und die berufliche Leistung ausschlaggebend sind (zusammen bilden diese beiden Aspekte den subjektiven und den objektiven Berufserfolg ab). Die Befunde gelten für diverse Berufe ebenso wie für den der Lehrerin und des Lehrers (vgl. Mayr 1994, 2003). Darüber hinaus weisen Forschungsstudien die Bedeutsamkeit personaler Prädispositionen entlang der drei genannten BigFive-Dimensionen (*Extraversion*, *Resilienz*, *Gewissenhaftigkeit*) schon für die Qualität des Lernens (gemäß Nutzung der Angebote, Anwendung von Lernstrategien und Ausdauer) im Studium nach – so auch im Lehramtsstudium (Mayr 2014, Baumert/Kunter 2010, Lipowsky 2010). Ein Mindestmaß dieser Dimensionen ist demnach nötig für einen erfolgreichen Abschluss des Lehramtsstudiums, was in die Empfehlung mündet, Persönlichkeitsmerkmale bei der individuellen Entscheidung für oder gegen ein Lehramtsstudium zu berücksichtigen (Mayr 2000, 2003, siehe auch Kap. 2.2.4). Diesen Befunden, denen zufolge die Dimensionen *Extraversion*, *Resilienz* und *Gewissenhaftigkeit*, die im Überblick über mögliche, persönliche Faktoren als Prädiktoren für den Studienerfolg stets herausragen, stehen andere Befunde gegenüber, denen zufolge primär die *Resilienz* prädiktiv sowohl für objektiven als auch subjektiven Berufserfolg bzw. -misserfolg ist (Boudreau et al. 2001, Judge et al. 2002, Barrick et al. 2001). Diesem allerdings stehen neuere Befunde entgegen, denen zufolge die *Gewissenhaftigkeit* keine signifikanten Effekte aufweist (vgl. Hülshager/Specht/Spinath 2006). Die Studie von Bipp (2007) weist wiederum bei einfachen Leistungsaufgaben eine Wirkung der beiden Dimensionen *Gewissenhaftigkeit* und *Resilienz* auch in Bezug auf Handlungsprozesse nach. Sie schließt daraus, dass Eigenschaften doch über den Handlungsprozess vermittelt in die Arbeitsleistung einfließen. Mayr (2014) zieht den Schluss, dass sich die in Bezug auf den Studienerfolg von Lehramtsanwärtern als bedeutsam herausgestellten Dimensionen *Gewissenhaftigkeit*, *Extraversion* und *Resilienz* ebenso in Praxisleistungen im Sinne guter Unterrichtsgestaltung niederschlagen. Inwiefern dies bei Lehrerinnen und Lehrern im Dienst tatsächlich der Fall ist, ist bislang allerdings nicht nachgewie-

sen¹⁵⁹; zumindest nicht, wenn die entsprechende Praxisleistung als hohe Unterrichtsqualität und nicht als hohe ökonomische Qualität (bspw. in Form weniger Krankentage) definiert wird. Vor diesem Hintergrund erscheint die Selektion angehender Lehrkräfte entlang ihrer (selbstwahrgenommenen) Persönlichkeit als ökonomische Maßnahme, die sich auf den Studienerfolg oder die berufliche Zufriedenheit bezieht. Über ökonomische Überlegungen hinaus erscheint eine derartige Selektion wenig fundiert. Zwar gelten Persönlichkeitskalen mittlerweile als valide Prädiktoren für die berufliche Leistung (Hülshager/Speccht/Spinath 2006), doch bleibt fraglich, ob der objektive Berufserfolg der untersuchten Berufsgruppen, die sich im Übrigen überwiegend auf den Finanzdienstleistungssektor beschränken, mit dem Berufserfolg von Lehrerinnen und Lehrern im Sinne von hoher Unterrichtsqualität mit einhergehenden Lernerfolgen und Lernleistungen ihrer Schülerinnen und Schüler vergleichbar ist. Daher wird an dieser Stelle ausdrücklich Mayrs Ansatz gefolgt, wenn er „Persönlichkeit eine wichtige Determinante der pädagogischen Kompetenz“¹⁶⁰ (Mayr 2003, S. 90) nennt und daraus die Notwendigkeit persönlichkeitsfördernder Bildungsangebote ableitet. Zwar leitet er aus seiner neueren Studie (Mayr/Neuweg 2018), die den Zusammenhang von unterrichtsmethodischer Grundeinstellung der Lehrkraft mit (selbsteingeschätzten) Persönlichkeitsmerkmalen der Lehrkraft (Alter und Big-Five-Dimensionen) untersucht, die Hypothese ab, dass die Art des Unterrichts aufgrund der Stabilität von Persönlichkeitsmerkmalen nicht wesentlich durch universitäre Ausbildung beeinflussbar ist. Allerdings sprechen der Ansatz und die Ergebnisse der vorliegenden Untersuchung deutlich für eine Kombination des hier als ökonomisch aufgefassten Aspekts der Selektion geeigneter Lehramtsstudierender anhand personaler Dispositionen mit einem sich in der weiteren Ausbildung verändernden Fokus, der sich auf das Trainieren (oder auch Entfalten) persönlicher Verhaltensweisen, die sich als förderlich in Hinblick auf die Unterrichtsqualität erwiesen haben, richtet. So ließen sich Ansätze des Dr.Fox-

¹⁵⁹ Erste Hinweise darauf, dass dies möglicherweise in Bezug auf die Dimensionen *Extraversion* und *Gewissenhaftigkeit* der Fall ist, liefern Baier et al. (2018). Allerdings sind die Effekte zum einen sehr schwach ausgeprägt und zum anderen wurden designbedingt lediglich die BigFive-Dimensionen *Extraversion* und *Gewissenhaftigkeit* in die auf das Fach Mathematik beschränkte Untersuchung einbezogen.

¹⁶⁰ Pädagogische Kompetenz wird in dieser Arbeit im Sinne von *Unterrichtsqualität* verstanden.

Effekts, der vermuten lässt, dass Schülereinschätzungen von persönlichen Eigenschaften ihrer Lehrerinnen und Lehrer verzerrt werden¹⁶¹ (vgl. Hofer 1981, Murray/Lawrence 1980), positiv nutzen. Das Gegenargument, dass sich diese Täuschung nicht über eine einmalige Lehrsituation hinaus aufrechterhalten ließe, da die Schülerinnen und Schüler im Schulalltag nicht dauerhaft unter dem blinderischen Einfluss verweilen, sondern ihre Lehrkräfte relativ gut im Laufe des Schuljahres in vielen Stunden gemeinsamen Unterrichts kennenlernen (Ditton 2002a), lässt sich angesichts der vorliegenden Ergebnisse nicht eindeutig halten, da nicht deutlich wird, inwiefern Lehrerinnen und Lehrer im Unterricht bewusst oder unbewusst eine Maske dauerhaft zur Schau tragen (so gesehen könnten Regressionen auch als Bias aufgefasst werden).

Bislang lauern in Bezug auf die Prognose der Bewährung im Lehramtsstudium zwei Gefahren: zum einen die der *Self-fulfilling Prophecy*, womit die freie Entfaltung der Lehramtsstudierenden gehemmt würde; zum anderen die der politischen Steuerung, wodurch die Anpasstheit an normierte Voraussetzungen bei der Lehrerausbildung mittels Selektion kontrolliert würde. Außerdem sehen Mayr und Neuweg (2006) in einer ausgeprägten Stabilitätsannahme von Persönlichkeit die Gefahr der Selektion zwischen Geeigneten und Ungeeigneten ohne Beachtung möglichen Entwicklungspotenzials. Allerdings sehen sie ebenso eine Gefahr in der Annahme einer veränderbaren Persönlichkeit: Sie könnte unter der Auffassung der beliebigen Modifizierbarkeit von Persönlichkeitsmerkmalen dann nicht Eingangskriterium, sondern Zielgröße darstellen (vgl. Teml/Unterweger 2002). Es ließen sich demgegenüber aber auch Chancen im Sinne der Formulierung von Entwicklungszielen und der Vervielfältigung von Lernarrangements ausmachen (Nieskens 2002), je nachdem, ob deterministischen oder probabilistischen Annahmen gefolgt wird. Der Mittelweg sähe nach Mayr und Neuweg (2006) demnach so aus, dass, ohne Persönlichkeitsveränderungen auszuschließen, der Fokus auf den bewussten Umgang mit Persönlichkeit durch individuelle Selbsterkenntnis (Umgang mit Stärken und Schwächen) gelegt würde, etwa in Form personenbezogener Reflexion, institutioneller Differenzierung und Individualisierung der Lernangebote.

¹⁶¹ Lehrkräfte, die auf bestimmte Verhaltensweisen trainiert wurden wie bspw. stimmliche Variationen und ausdrucksvolle Bewegungen, erhalten demnach bessere Ratings als eine Kontrollgruppe.

Es mehren sich in der Lehrerausbildung Ansätze (bspw. PARCOUR der Universität Passau¹⁶², TESAT in Österreich (Neubauer/Koschmieder/Krammer/Mayr et al. 2017)), die Persönlichkeitsinventare in ihre Selektionsverfahren für Lehramtsanwärter aufzunehmen gedenken – ausnahmslos aus Perspektive des Selbstberichts heraus –, um geeignete Kandidaten von weniger oder ungeeigneten zu unterscheiden. Eine einseitige, ökonomische Interpretation dieser Selektion birgt allerdings die Gefahr der Einschränkung der Vielfalt an Persönlichkeiten in Schulen, vor allem sofern nicht entgegenwirkend während des Lehramtsstudiums – angesichts der Relevanz, die Persönlichkeit für die Unterrichtsqualität hat – die Notwendigkeit heterogener personaler Facetten vermittelt wird. Ähnlich Schauspielern könnten Lehrerinnen und Lehrer jedoch auch lernen, bestimmte Eigenschaften einzusetzen, um die Wahrnehmung der Prozessqualität des Unterrichts aus Schülersicht positiv zu beeinflussen. Konkret postulieren dementsprechend Bürmann und Bürmann (2003) bspw. die Gestalttherapie als berufsbegleitende Fortbildung für Lehrerinnen und Lehrer; zum einen zur Arbeit an der eigenen Persönlichkeit und zum anderen zur Kultivierung einer Wahrnehmungsoffenheit und einer wechselseitigen Beziehungsstruktur.

„Etwas pointiert könnte man sagen: auf das ‚objektive‘ Lehrerverhalten kommt es im Unterricht gar nicht an, sondern darauf, wie es subjektiv von den Schülern wahrgenommen wird (Steltmann 1992, S. 571). Insofern würden die Schülerbewertungen das Wesentliche am Unterricht spiegeln.“ (Terhart 1997, S. 55)

¹⁶² „Derzeit ist PARCOURS ein Beratungsangebot. Sie behalten unabhängig von den Ergebnissen in jedem Fall Ihren Studienplatz.“ Näheres unter <http://www.phil.uni-passau.de/schulpaedagogik/forschungprofilelemente/parcours/ueber-parcours/>

20 Limitationen und Ausblick

Die vorliegende Arbeit unterliegt selbstverständlich gewissen Limitationen, die gleichsam einen Ausblick für weitere wünschenswerte Forschungen bieten, um die in dieser Arbeit ermittelten Ergebnisse zu ergänzen oder gegebenenfalls zu modifizieren: So wurden lediglich Querschnittsdaten verwendet und damit ein einmaliger Einblick in eine bestimmte Konstellation zu einem bestimmten Zeitpunkt gewonnen – wenn auch der bewusst gewählte Zeitpunkt zum Ende des Schuljahres als günstig gelten kann. Längsschnittuntersuchungen hingegen erst können ermitteln, inwiefern die vorliegenden Befunde als robust gelten können. Vor allem unter Einbezug von Lernstandsmessungen ließe sich ein wesentlich differenzierterer Blick auf die Effekte eröffnen.

Wie bereits vorab erwähnt (vgl. Kap. 11.4 und 17.2) wurden die Lehrerselbsteinschätzungen ohne den kontextualen Bezug zur Schule erfasst. Hier wäre eine Anpassung des eingesetzten Instruments an die Lehrerrolle wünschenswert, etwa durch die Einleitung der Persönlichkeitsselbsteinschätzungen mit „Ich bin ein/e Lehrer/in, die/der... „, oder „Als Lehrer/in bin ich ein Mensch, der...“. Damit ließe sich überprüfen, ob eine kontextspezifische Selbsteinschätzung möglicherweise deutlicher mit den Schülereinschätzungen übereinstimmt, als dies in der vorliegenden Untersuchung der Fall war.

Die Arbeit ist auf die Untersuchung der Fächer Deutsch und Mathematik begrenzt, wobei anzunehmen ist, dass sich die erforschten Effekte in anderen Fächern wiederum in anderen Nuancen abzeichnen und die Übertragung der Ergebnisse auf andere Fächer daher vorsichtig erfolgen sollte. Die Untersuchung weiterer Unterrichtsfächer wäre unbedingt wünschenswert, um einen umfassenden und differenzierten Einblick in die Bedeutung der Lehrerpersönlichkeit für den schulischen Unterricht zu gewinnen.

Interessant wäre außerdem, in welchem Ausmaß die verschiedenen Prozessmerkmale je nach Fach relevant für die empfundene Unterrichtsqualität sind, um zu ermessen, in welchem Ausmaß dies die einzelnen Persönlichkeitsdimensionen der Lehrkraft insgesamt sind. Dies bedürfte weiterer differenzierter Untersuchungen. Zudem sind die erfassten Lehrpersonen nur mit einer Klasse und in jeweils einem Fach repräsentiert, d.h. dass eine möglicherweise bestehende intrapersonale Varianz je nach Klasse oder Fach nicht untersucht werden konnte.

Erst mit der Untersuchung einzelner sowohl fachspezifisch als auch fachübergreifend durch mehrere Klassen eingeschätzter Lehrpersonen lassen sich die vorliegenden Befunde sicher verifizieren.

In dieser Arbeit wurde generell die Regression der Unterrichtsqualität auf die Lehrerpersönlichkeit angenommen und verfolgt. Eine umgekehrte Wirkrichtung könnte jedoch bezüglich der Schülerperspektive ebenfalls plausibel hergeleitet werden (vgl. auch Baier et al. 2018). In diesem Zusammenhang wäre des Weiteren zu überprüfen, ob und inwiefern Schülerinnen und Schüler tatsächlich zwischen den BigFive-Dimensionen der Lehrkräfte und Prozessmerkmalen des Unterrichts differenzieren.

Weitere Forschung mit der Ausrichtung auf den in der vorliegenden Arbeit gewählten Ansatz wäre demnach ebenso wünschenswert wie angebracht, um die berichteten Ergebnisse weiter zu präzisieren und zu verifizieren. Dabei wäre auch die Fokussierung auf geschlechtsspezifische Aspekte, die hier trotz Hinweisen auf deren Relevanz wenig Berücksichtigung finden, angebracht. Ebenfalls wurden altersspezifische Aspekte, zu denen einschlägige Forschungsbefunde bestehen, gänzlich außen vor gelassen.

Literaturverzeichnis

- Abele, A. E. (2014). Akteur-Beobachter-Unterschied. In M. A. Wirtz (Hrsg.), *Dorsch. Lexikon der Psychologie* (17. Aufl.), 114. Bern: Hans Huber.
- Abele, A. E. & Wojciszke, B. (2007). Agency and communion from the perspective of self versus others. *Journal of Personality and Social Psychology*, 93 (5), 751–763. <https://doi.org/10.1037/0022-3514.93.5.751>
- Abrami, P. C., Leventhal, L. & Perry, R. P. (1982). Educational Seduction. *Review of Educational Research*, 52 (3), 446–464.
- Aleamoni, L. M. (1981). Student ratings of instruction. In J. Millman (Ed.), *Handbook of teacher evaluation*, 110–145. Beverly Hills: Sage.
- Aleamoni, L. M. (1999). Student Rating Myths Versus Research Facts from 1924 to 1998. *Journal of Personnel Evaluation in Education*, 13 (2), 153–166. <https://doi.org/10.1023/A:1008168421283>
- Allemann-Ghionda, C. & Terhart, E. (Hrsg.). (2006). *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern. Ausbildung und Beruf* (Zeitschrift für Pädagogik. Beihefte, Bd. 51). Weinheim: Beltz.
- Allison, P. (2012). *When can you safely ignore multicollinearity?* Zugriff am 27.07.17. Verfügbar unter <http://statisticalhorizons.com/multicollinearity>
- Allport, G. W. (1937). *Personality. A psychological interpretation*. New York: Holt Rinehart and Winston.
- Allport, G. W. & Odbert, H. S. (1936). Trait-names. A psycho-lexical study. *Psychological Monographs*, 47 (1), i-171. <https://doi.org/10.1037/h0093360>
- Allport, G. W. (1961). *Pattern and growth in personality* (Psychology). New York: Holt Rinehart & Winston.
- Altwicker, T. (1999). *Zu den linguistischen Grundlagen der Diskursethik bei Jürgen Habermas*. Zugriff am 01.08.2008. Verfügbar unter <http://www.gs.uni-heidelberg.de/sprache2/altwick1.htm>
- Apel, H.-J. (2009). Klassenführung. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (UTB Schulpädagogik, Pädagogik, Bd. 8423, 2., aktualisierte Aufl.), 171–175. Bad Heilbrunn: Klinkhardt.
- Arnold, K.-H. (Hrsg.). (2007). *Unterrichtsqualität und Fachdidaktik*. Bad Heilbrunn: Klinkhardt.
- Arnold, K.-H., Sandfuchs, U. & Wiechmann, J. (Hrsg.). (2009). *Handbuch Unterricht* (UTB Schulpädagogik, Pädagogik, Bd. 8423, 2., aktualisierte Aufl.). Bad Heilbrunn: Klinkhardt.

- Asendorpf, J. (2014). Persönlichkeit. In M. A. Wirtz (Hrsg.), *Dorsch. Lexikon der Psychologie* (17. Aufl.), 1244–1245. Bern: Hans Huber.
- Asendorpf, J. & Neyer, F. (2012). *Psychologie der Persönlichkeit* (5. Aufl.). Berlin: Springer.
- Asendorpf, J. B. (2004). *Psychologie der Persönlichkeit* (Springer-Lehrbuch, 3., überarbeitete und aktualisierte Auflage). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-09570-6>
- Babu, S. & Medro, R. (2003). *Teacher Accountability. HLM-based teacher effectiveness indices in the investigation of teacher effects on student achievement in a state assessment program*. Annual Meeting of the American Educational Research Association, Chicago, Illinois, USA.
- Back, M. D., Baumert, A., Denissen, Jaap J. A., Hartung, F.-M., Penke, L., Schmukle, S. C. et al. (2011). PERSOC: A Unified Framework for Understanding the Dynamic Interplay of Personality and Social Relationships. *European Journal of Personality*, 25, 90–107.
- Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B. et al. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological science*, 21 (3), 372–374. <https://doi.org/10.1177/0956797609360756>
- Back, M. D. & Vazire, S. (2012). Knowing Our Personality. In S. Vazire & T. D. Wilson (Eds.), *Handbook of self-knowledge*, 131–149. New York: Guilford Press.
- Backhaus, K. (2004). *Persönlichkeit als Forschungsgegenstand der Psychologie. Eine Einführung in das Big Five-Persönlichkeitsmodell*. Verfügbar unter <http://www.psyreon.de/content/e479/e480/Publikationen/persoenlichkeit.pdf>
- Baier, F., Decker, A.-T., Voss, T., Kleickmann, T., Klusmann, U. & Kunter, M. (2018). What makes a good teacher? The relative importance of mathematics teachers' cognitive ability, personality, knowledge, beliefs, and motivation for instructional quality. *British Journal of Educational Psychology*, 1-20. <https://doi.org/10.1111/bjep.12256>
- Barrick, M. R. & Mount, M. K. (1991). The Big Five Personality Dimensions And Job Performance. A Meta-Analysis. *Personnel Psychology*, 44 (1), 1–26. <https://doi.org/10.1111/j.1744-6570.1991.tb00688.x>
- Barrick, M. R., Mount, M. K. & Judge, T. A. (2001). Personality and Performance at the Beginning of the New Millennium. What Do We Know and Where Do We Go Next? *International Journal of Selection and Assessment*, 9 (1&2), 9–30. <https://doi.org/10.1111/1468-2389.00160>
- Bartz, C., Jäger, L., Krause, M. & Linz, E. (Hrsg.). (2012). *Handbuch der Mediologie. Signaturen des Medialen*. München: Fink.

- Baumeister, R. F. & Cairns, K. J. (1992). Repression and self-presentation. When audiences interfere with self-deceptive strategies. *Journal of Personality and Social Psychology*, 62 (5), 851–862. <https://doi.org/10.1037/0022-3514.62.5.851>
- Baumeister, R. F., Vohs, K. D. & Funder, D. C. (2007). Psychology as the Science of Self-Reports and Finger Movements: Whatever Happened to Actual Behavior? *Perspectives on psychological science: a journal of the Association for Psychological Science*, 2 (4), 396–403. <https://doi.org/10.1111/j.1745-6916.2007.00051.x>
- Baumert, J. & Köller, O. (1998). Interest research in secondary level I: An overview. 241-256. In L. Hoffmann (Ed.), *Interest and learning. Proceedings of the Seeon Conference on Interest and Gender* (IPN, vol. 164). Kiel: IPN.
- Baumert, J., Roeder, P. M., Gruehn, S., Heyn, S. & Köller, O. u. (1996). Bildungsverläufe und psychosoziale Entwicklung im Jugendalter (BIJU). In Treumann K. P., Neubauer G., Möller R. & Abel J. (Hrsg.), *Methoden und Anwendungen empirischer pädagogischer Forschung*, 170–180. Münster/New York: Waxmann.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B, Clausen, M., Hosenfeld, I. & Neubrand, J. (1997). *TIMSS. Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde*. Opladen: Leske + Budrich.
- Beaton, A. E. (1996). *Mathematics achievement in the middle school years. IEA's third international mathematics and science study (TMSS); November 1996*. Chestnut Hill, Mass.: TIMSS Internat. Study Center Boston College.
- Beauducel, A. & Kersting, M. (2010). *START-P. Testbatterie für Berufseinsteiger-Persönlichkeit*. Göttingen: Hogrefe.
- Bendig, A. W. (1953). The Relation of Level of Course Achievement to Students' Instructor and Course Ratings in Introductory Psychology. *Educational and Psychological Measurement*, 13 (3), 437–448. <https://doi.org/10.1177/001316445301300307>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107 (2), 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Bergsdorf, W. (Hrsg.). (2002). *Herausforderungen der Bildungsgesellschaft*. Weimar: Rhino-Verlag.
- Berkemeyer, N., Bos, W., Holtappels, H. G., Meetz, F. & Rollett, W. (2010). „Ganz In“: Das Ganztagsgymnasium in Nordrhein-Westfalen–Bestandsaufnahme und Perspektiven eines Schulentwicklungsprojekts. *Jahrbuch der Schulentwicklung*, 16, 131–152.
- Bettencourt, E. M., Gillett, M. H., Gall, M. D. & Hull, R. E. (1983). Effects of Teacher Enthusiasm Training on Student On-task Behavior and Achievement. *American Educational Research Journal*, 20 (3), 435–450. <https://doi.org/10.3102/00028312020003435>

- Biddle, B. J., Good, T. L. & Goodson, I. F. (Eds.). (1997). *International Handbook of Teachers and Teaching* (Springer International Handbooks of Education, vol. 3). Dordrecht: Springer Netherlands. <https://doi.org/10.1007/978-94-011-4942-6>
- Bierhoff, H.-W. & Frey, D. (Hrsg.). (2006). *Handbuch der Sozialpsychologie und Kommunikationspsychologie*. Göttingen: Hogrefe Verlag.
- Bierhoff, H.-W. & Frey, D. (Hrsg.). (2006). *Handbuch der Sozialpsychologie und Kommunikationspsychologie*. Göttingen: Hogrefe Verlag.
- Bipp, T. (2006). *Persönlichkeit - Ziele - Leistung. Der Einfluss der Big Five Persönlichkeitseigenschaften auf das zielbezogene Leistungshandeln*. Dissertation.
- Blocker, C., Conway, J., Demortier, L., Heinrich, J., Junk, T., Lyons, L. et al. (2006). Simple Facts about P-Values. *CDF/MEMO/STATISTICS/PUBLIC, 2006* (8023). CDF Statistics Committee.
- Blömeke, S. (2009). Voraussetzungen bei der Lehrperson. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht* (UTB Schulpädagogik, Pädagogik, Bd. 8423, 2., aktualisierte Aufl.), 122–127. Bad Heilbrunn: Klinkhardt.
- Blömeke, S., Reinhold, P., Tulodziecki, G. & Wildt, J. (Hrsg.). (2004). *Handbuch Lehrerbildung*. Düsseldorf: Klinkhardt.
- Blunt, A. (1991). The Effects of Anonymity and Manipulated Grades on Student Ratings of Instructors. *Community College Review, 18* (4), 48–54. <https://doi.org/10.1177/009155219101800409>
- Bohnsack, F. (2004). Persönlichkeitsbildung von Lehrerinnen und Lehrern. In S. Blömeke, P. Reinhold, G. Tulodziecki & J. Wildt (Hrsg.), *Handbuch Lehrerbildung*, 152–164. Düsseldorf: Klinkhardt.
- Bollen, K. A. & Long, J. S. (Eds.). (1993). *Testing structural equation models* (Sage focus editions, vol. 154, [Nachdr.]. Newbury Park, Calif.: Sage.
- Borkenau, P. & Ostendorf, F. (1993). *NEO-Fünf-Faktoren-Inventar (NEO-FFI) nach Costa und McCrae*. Göttingen: Hogrefe.
- Borkenau, P. & Ostendorf, F. (2001). NEO-Fünf-Faktoren-Inventar. In W. Sarges & H. Wottawa (Hrsg.), *Handbuch wirtschaftspsychologischer Testverfahren*. Lengerich: Pabst Science Publishers.
- Bortz, J. & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-12770-0>
- Boudreau, J. W., Boswell, W. R. & Judge, T. A. (2001). Effects of Personality on Executive Career Success in the United States and Europe. *Journal of Vocational Behavior, 58* (1), 53–81. <https://doi.org/10.1006/jvbe.2000.1755>

- Brandenburg, D. C., Slinde, J. A. & Batista, E. E. (1977). Student ratings of instruction. Validity and normative interpretations. *Research in Higher Education*, 7 (1), 67–78. <https://doi.org/10.1007/BF00991945>
- Brigham, F. J. (1991). *Generating Excitement: Teacher Enthusiasm and Students with Learning Disabilities*. Paper presented at the Annual Meeting of the Council for Learning Disabilities, Minneapolis, MN.
- Bromme, R. (1997). Kompetenzen, Funktionen und unterrichtliches Handeln des Lehrers. *Psychologie des Unterrichts und der Schule*, 3, 177–212.
- Bromme, R. & Haag, L. (2008). "Forschung zur Lehrerpersönlichkeit". In W. Helsper & J. Böhme (Hrsg.), *Handbuch der Schulforschung* (2., durchgesehene und erweiterte Auflage), 803–819. Wiesbaden: VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH Wiesbaden.
- Brophy, J. (2000). *Teaching. Educational Practices Series--1*. Zugriff am 23.02.2017. Verfügbar unter <https://files.eric.ed.gov/fulltext/ED440066.pdf>
- Brosius, F. (2008). *SPSS-Programmierung. Effizientes Datenmanagement und Automatisierung mit SPSS-Syntax; [Syntaxbeispiele zum Download. Einführung in die Programmierung mit SPSS-Syntax, Dateimanagement und Datenaustausch. Berechnung, Aufbereitung und Analyse von Daten. Makro-Programmierung und Automatisierung von Programmabläufen]* (2. Aufl.). Heidelberg: mitp Redline.
- Brosius, F. (2013). *SPSS 21. [fundierte Einführung in SPSS und in die Statistik; alle statistischen Verfahren mit praxisnahen Beispielen; inklusive CD-ROM]* (1. Aufl.). Heidelberg: mitp Verl.-Gruppe Hüthig Jehle Rehm.
- Browne, M. W. & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models, Vol. 154*, 136–162. Newbury Park, Calif.: Sage.
- Bürmann, I. & Bürmann, J. (2003). Gestaltpädagogik als phänomenologische und tiefenpsychologische Fundierung von Lehrerhandeln. In F. C. Sauter (Hrsg.), *Schulwirklichkeit und Wissenschaft. Ausgewählte Kongressbeiträge von Didaktikern, Pädagogen, Psychologen* (Schriftenreihe Erziehung - Unterricht - Bildung, Bd. 101), 223–241. Hamburg: Kovač.
- Campbell, R. J. (2004). *Assessing teacher effectiveness. Developing a differentiated model*. London: RoutledgeFalmer.
- Carroll, J. B. (1963). A model of school learning. *Teachers College Record*, 64, 723–733.
- Carroll, J. B. (1973). Ein Modell schulischen Lernens. In W. Edelstein & D. Hopf (Hrsg.), *Bedingungen des Bildungsprozesses. Psychologische und pädagogische Forschungen zum Lehren und Lernen in der Schule* (Konzepte der Humanwissenschaften), 234–250. Stuttgart: Klett.

- Carroll, J. B. (1989). The Carroll Model. A 25-year retrospective and prospective view. *Educational Researcher*, 18 (1), 26–31.
- Caspi, A. & Roberts, B. W. (2001). Personality Development Across the Life Course. The Argument for Change and Continuity. *Psychological Inquiry*, 12 (2), 49–66. https://doi.org/10.1207/S15327965PLI1202_01
- Centra, J. A. (1977). Student Ratings of Instruction and Their Relationship to Student Learning. *American Educational Research Journal*, 14 (1), 17–24. <https://doi.org/10.3102/00028312014001017>
- Centra, J. A. (1974). The Relationship Between Student and Alumni Ratings of Teachers. *Educational and Psychological Measurement*, 34 (2), 321–325. <https://doi.org/10.1177/001316447403400212>
- Centra, J. A. (1993). *Reflective faculty evaluation. Enhancing teaching and determining faculty effectiveness* (The Jossey-Bass higher and adult education series, 1st ed.). San Francisco: Jossey-Bass.
- Christ, O. & Schlüter, E. (2012). *Strukturgleichungsmodelle mit Mplus. Eine praktische Einführung*, Sozialwissenschaften 10-2012. München: Oldenbourg.
- Clausen, M. (2002). *Unterrichtsqualität: eine Frage der Perspektive? Empirische Analysen zur Übereinstimmung, Konstrukt- und Kriteriumsvalidität*, Pädagogische Psychologie und Entwicklungspsychologie, Bd. 29. Münster: Waxmann.
- Clausen, M., Reusser, K. & Klieme, E. (2003). Unterrichtsqualität auf der Basis hoch-inferenter Unterrichtsbeurteilungen. Ein Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31 (2), 122–141.
- Clayson, D. E. & Haley, D. A. (1990). Student Evaluations in Marketing. What is Actually being Measured? *Journal of Marketing Education*, 12 (3), 9-17. <https://doi.org/10.1177/027347539001200302>
- Clayson, D. E. & Sheffet, M. J. (2006). Personality and the Student Evaluation of Teaching. *Journal of Marketing Education*, 28, 149–160.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement. A meta-analysis of multisection validity studies. *Review of Educational Research*, 51 (3), 281–309.
- Cohen, P. A. (1980). Effectiveness of student-rating feedback for improving college instruction. A meta-analysis of findings. *Research in Higher Education*, 13 (4), 321–341. <https://doi.org/10.1007/BF00976252>
- Collins, M. L. (1978). Effects of Enthusiasm Training on Preservice Elementary Teachers. *Journal of Teacher Education*, 29 (1), 53–57. <https://doi.org/10.1177/002248717802900120>

- Connely, B.S. & Ones, D.S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, 136, 1092-1122.
- Costa, P. T. & McCrae, R. R. (1980). Influence of extraversion and neuroticism on subjective well-being. Happy and unhappy people. *Journal of Personality and Social Psychology*, 38 (4), 668–678. <https://doi.org/10.1037/0022-3514.38.4.668>
- Costa, P. T. & McCrae, R. R. (1988). Personality in adulthood. A six-year longitudinal study of self-reports and spouse ratings on the NEO Personality Inventory. *Journal of Personality and Social Psychology*, 54 (5), 853–863. <https://doi.org/10.1037/0022-3514.54.5.853>
- Costa, P. T. & McCrae, R. R. (Hrsg.). (1989) NEO PI/FFI Manual Supplement [Themenheft]. Odessa, Florida: Psychological Assessment Resources.
- Costin, F. & Grush, J. E. (1973). Personality correlates of teacher-student behavior in the college classroom. *Journal of Educational Psychology*, 65 (1), 35–44. <https://doi.org/10.1037/h0034814>
- Craft, A. (Ed.). (1992). *Quality assurance in higher education. Proceedings of an international conference Hong Kong, 1991*. London: Falmer Press.
- Cuntz, M. & Krause, M. (2012). (Hyper-)Normalisierung. In C. Bartz, L. Jäger, M. Krause & E. Linz (Hrsg.), *Handbuch der Mediologie. Signaturen des Medialen*, 191–201. München: Fink.
- Daniels, Z. (2008). *Entwicklung schulischer Interessen im Jugendalter*, Pädagogische Psychologie und Entwicklungspsychologie, Bd. 69. Zugl.: Potsdam, Berlin, Univ., Diss., 2004. Münster: Waxmann.
- Deci, E. L. & Ryan, R. M. (1985). The general causality orientations scale. Self-determination in personality. *Journal of Research in Personality*, 19 (2), 109–134. [https://doi.org/10.1016/0092-6566\(85\)90023-6](https://doi.org/10.1016/0092-6566(85)90023-6)
- Deci, E. L. & Ryan, R. M. (1993). Die Selbstbestimmungstheorie der Motivation und ihre Bedeutung für die Pädagogik. *Zeitschrift für Pädagogik*, 39 (2), 223–238.
- Dehne, M. & Schupp, J. (2007). *Persönlichkeitsmerkmale im Sozio-oekonomischen Panel (SOEP). Konzept, Umsetzung und empirische Eigenschaften*. Berlin: Deutsches Institut für Wirtschaftsforschung.
- DePaulo, B. M. (1992). Nonverbal behavior and self-presentation. *Psychological Bulletin*, 111 (2), 203–243. <https://doi.org/10.1037/0033-2909.111.2.203>
- Deshpande, A. S., Webb, S. C. & Marks, E. (1970). Student Perceptions of Engineering Instructor Behaviors and Their Relationships to the Evaluation of Instructors and Courses. *American Educational Research Journal*, 7 (3), 289–305. <https://doi.org/10.3102/00028312007003289>

- Ditton, H. (2002a). Lehrkräfte und Unterricht aus Schülersicht. Ergebnisse einer Untersuchung im Fach Mathematik. *Zeitschrift für Pädagogik* (2), 262–286.
- Ditton, H. (2000). Qualitätskontrolle und Qualitätssicherung in Schule und Unterricht. Ein Überblick zum Stand der empirischen Forschung. *Zeitschrift für Pädagogik*, 41, 73–92.
- Ditton, H. (2002). Unterrichtsqualität. Konzeptionen, methodische Überlegungen und Perspektiven. *Unterrichtswissenschaft*, 20 (3), 197–212.
- Ditton, H. & Arnold, B. (2004). Wirksamkeit von Schülerfeedback zum Fachunterricht. In J. Doll & M. Prenzel (Hrsg.), *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung*, 152–170. Waxmann Verlag.
- Doll, J. & Prenzel, M. (Hrsg.). (2004). *Bildungsqualität von Schule. Lehrerprofessionalisierung, Unterrichtsentwicklung und Schülerförderung als Strategien der Qualitätsverbesserung*. Münster: Waxmann Verlag.
- Döring, K. W. (1992). *Lehrerverhalten. Forschung - Theorie - Praxis; ein Lehrbuch*. Weinheim u.a.: Beltz.
- Dunkin, M. J. & Biddle, B. J. (1974). *The study of teaching*. New York: Holt Rinehart and Winston.
- Dunnette, M. D. & Hough, L. M. (Eds.). (1991). *Handbook of industrial and organizational psychology* (2nd ed.). Palo Alto/Calif.: Consulting Psychologists Press.
- Dworkin, A. G. (1997). Coping with Reform: The Intermix of Teacher Morale, Teacher Burn-out and Teacher Accountability. In B. J. Biddle, T. L. Good & I. F. Goodson (Eds.), *International Handbook of Teachers and Teaching* (Springer International Handbooks of Education, vol. 3), 459–498. Dordrecht: Springer Netherlands.
- Edelstein, W. & Hopf, D. (Hrsg.). (1973). *Bedingungen des Bildungsprozesses. Psychologische und pädagogische Forschungen zum Lehren und Lernen in der Schule* (Konzepte der Humanwissenschaften). Stuttgart: Klett.
- Eder, F. (1985). *Schul- und Lernprobleme am BORG. Ein Vergleich von Schülern der Langform und der Oberstufenform des Gymnasiums.*, Pädagogisches Institut des Bundes Oberösterreich, Abteilung allgemeinbildende höhere Schulen. Linz: Pädagogisches Institut des Bundes.
- Eder, F. & Bergmann, C. (2004). Der Einfluss von Interessen auf die Lehrer-Wahrnehmung von Schülerinnen und Schülern. *Empirische Pädagogik*, 18 (4), 410–431.
- Ehlert, H., Welbers, U. & Eckermann, V. K. (Hrsg.). (1999). *Handbuch Praxisinitiativen an Hochschulen. Berufsorientierende Angebote für Studierende an Universitäten*, Hochschulwesen - Wissenschaft und Praxis. Neuwied: Luchterhand.

- Eid, M., Gollwitzer, M. & Schmitt, M. (2013). *Statistik und Forschungsmethoden. Lehrbuch* (3., korrigierte Aufl.). Weinheim: Beltz.
- Einsiedler, W. (1997). Unterrichtsqualität und Leistungsentwicklung: Literaturüberblick. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter*, 225–240. Weinheim: Beltz Psychologie-Verl.-Union.
- Einsiedler, W. & Treinies, G. (2006). Effects of Teaching Methods, Class Effects, and Patterns of Cognitive Teacher-Pupil Interactions in an Experimental Study in Primary School Classes. *School Effectiveness and School Improvement*, 8 (3), 327–353. <https://doi.org/10.1080/0924345970080303>
- Erdfelder, E., Faul, F., Buchner, A. & Cüpper, L. (2010). Effektgröße und Teststärke. In H. Holling (Hrsg.), *Handbuch Statistik, Methoden und Evaluation*, Handbuch der Psychologie, Bd. 13, 358–369. Göttingen: Hogrefe.
- Erdle, S., Murray, H. G. & Rushton, J. P. (1985). Personality, classroom behavior, and student ratings of college teaching effectiveness. A path analysis. *Journal of Educational Psychology*, 77 (4), 394–407.
- Erikson, E. H. (1966). *Identität und Lebenszyklus*. Berlin: Suhrkamp.
- Eysenck, H. J. (1947). *Dimensions of Personality*. London: Routledge & Kegan Paul.
- Eysenck, H. J. & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. San Diego, CA: Educational and Industrial Testing Service.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E. & Büttner, G. (2014). Grundschulunterricht aus Schüler-, Lehrer- und Beobachterperspektive. Zusammenhänge und Vorhersage von Lernerfolg. *Zeitschrift für Pädagogische Psychologie*, 28 (3), 127–137. <https://doi.org/10.1024/1010-0652/a000129>
- Feldman, K. A. (1976). Grades and college students' evaluations of their courses and teachers. *Research in Higher Education*, 4 (1), 69–111. <https://doi.org/10.1007/BF00991462>
- Feldman, K. A. (1977). Consistency and variability among college students in rating their teachers and courses: A review and analysis. *Higher Education*, 6 (3), 223–274. <https://doi.org/10.1007/BF00991288>
- Feldman, K. A. (1978). Course characteristics and college students' ratings of their teachers. What we know and what we don't. *Research in Higher Education*, 9 (3), 199–242. <https://doi.org/10.1007/BF00976997>
- Feldman, K. A. (1979). The significance of circumstances for college students' ratings of their teachers and courses. *Research in Higher Education*, 10 (2), 149–172.
- Feldman, K. A. (1986). The perceived instructional effectiveness of college teachers as related to their personality and attitudinal characteristics. A review and synthesis. *Higher Education*, 24 (2), 139–213. <https://doi.org/10.1007/BF00991885>

- Feldman, K. A. (1988). Effective college teaching from the students' and faculty's view. Matched or mismatched priorities? *Research in Higher Education*, 28 (4), 291–329. <https://doi.org/10.1007/BF01006402>
- Feldman, K. A. (1989). The association between student ratings of specific instructional dimensions and student achievement. Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, 30 (6), 583–645. <https://doi.org/10.1007/BF00992392>
- Feldman, K. A. (1993). College Students' Views of Male and Female College Teachers. Part II Evidence from Students' Evaluations of Their Classroom Teachers. *Research in Higher Education*, 34 (2), 151–211.
- Fend, H. (1977). *Schulklima. Soziale Beeinflussungsprozesse in der Schule*. Weinheim: Beltz.
- Fend, H. & Specht, W. (1986). *Erziehungsumwelten. Bericht aus dem Projekt "Entwicklung im Jugendalter"*. Sozialwissenschaftliche Fakultät, Universität Konstanz.
- Fondel, E., Lischetzke, T., Weis, S. & Gollwitzer, M. (2015). Zur Validität von studentischen Lehrveranstaltungsevaluationen. *Diagnostica*, 61 (3), 124–135. <https://doi.org/10.1026/0012-1924/a000141>
- Frey, A. & Asseburg, R. (Hrsg.). (2009). *PISA-2006-Skalenhandbuch. Dokumentation der Erhebungsinstrumente*. Münster: Waxmann.
- Frey, P. W. (1978). A two-dimensional analysis of student ratings of instruction. *Research in Higher Education*, 9 (1), 69–91. <https://doi.org/10.1007/BF00979187>
- Fromme, J. (1997). *Pädagogik als Sprachspiel. Zur Pluralisierung der Wissensformen im Zeichen der Postmoderne*, Pädagogik, Theorie und Praxis. Neuwied: Luchterhand.
- Funder, D. C. (1999). *Personality judgment. A realistic approach to person perception*. San Diego: Academic Press.
- Furnham, A. (1996). The big five versus the big four. The relationship between the Myers-Briggs Type Indicator (MBTI) and NEO-PI five factor model of personality. *Personality and Individual Differences*, 21 (2), 303–307. [https://doi.org/10.1016/0191-8869\(96\)00033-5](https://doi.org/10.1016/0191-8869(96)00033-5)
- Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, 23 (5), 369–401. <https://doi.org/10.1002/per.724>
- Gallrein, A.-M. B., Carlson, E. N., Holstein, M. & Leising, D. (2013). You spy with your little eye. People are “blind” to some of the ways in which they are consensually seen by others. *Journal of Research in Personality*, 47 (5), 464–471. <https://doi.org/10.1016/j.jrp.2013.04.001>
- Gardner, P. L. (1987). Schülerinteressen an Naturwissenschaften und Technik. In M. Lehrke & L. Hofmann (Hrsg.), *Schülerinteressen am naturwissenschaftlichen Unterricht. Untersuchungen und Erklärungen; Beiträge zum 12. IPN-Symposium; die vorl. Arbeit entstand am*

- Inst. für die Pädagogik der Naturwissenschaften (IPN) an der Univ. Kiel, Didaktik der Naturwissenschaften, Bd. 12, 13–38. Köln: Aulis-Verl. Deubner.*
- Geiser, C. (2013). *Data analysis with Mplus* (Methodology in the social sciences). New York: Guilford Press.
- Gentry, M., Gable, R. K. & Rizza, M. G. (2002). Students' Perceptions of Classroom Activities. Are There Grade-Level and Gender Differences? *Journal of Educational Psychology, 94* (3), 539–544.
- Gerhard, U. (1999). Borkenau, P. & Ostendorf, F. (1993). NEO-Fünf-Faktoren Inventar (NEO-FFI) nach Costa und McCrae. Göttingen. Hogrefe. *Zeitschrift für Klinische Psychologie und Psychotherapie, 28* (2), 145–146. <https://doi.org/10.1026//0084-5345.28.2.145>
- Goffman, E. (2016). *Stigma. Über Techniken der Bewältigung beschädigter Identität*, Suhrkamp-Taschenbuch Wissenschaft, Bd. 140, 23. Auflage. Frankfurt am Main: Suhrkamp.
- Goffman, E. & Kihm, A. (1975). *Stigmaté: les usages sociaux des handicaps*. Paris: Editions de minuit.
- Goldberg, L. R. (1990). An alternative "description of personality". The Big-Five factor structure. *Journal of Personality and Social Psychology, 59* (6), 1216–1229. <https://doi.org/10.1037/0022-3514.59.6.1216>
- Goldstein, E. B. (2011). *Cognitive psychology. Connecting mind, research and everyday experience*. Stamford, CT: Cengage Learning.
- Göllner, R., Wagner, W., Klieme, E., Lüdtke, O., Nagengast, B. & Trautwein, U. (2016). *Erfassung der Unterrichtsqualität mithilfe von Schülerurteilen: Chancen, Grenzen und Forschungsperspektiven*, Bildungsforschung, Band 44, Stand August 2016. Berlin: Bundesministerium für Bildung und Forschung (BMBF) Referat Bildungsforschung.
- Graham, J. W. (2003). Adding Missing-Data-Relevant Variables to FIML-Based Structural Equation Models. *Structural Equation Modeling: A Multidisciplinary Journal, 10* (1), 80–100. https://doi.org/10.1207/S15328007SEM1001_4
- Gräsel, C. & Göbel, K. (2015). Unterrichtsqualität. In H. Reinders (Hrsg.), *Empirische Bildungsforschung. Gegenstandsbereiche, 2.*, überarb. Aufl., 107–119. Wiesbaden: Springer VS.
- Greimel, B. (2002). Lehrerevaluation durch Beurteilungen der Lernenden. Eine Analyse des Standes der Evaluationsforschung. *Zeitschrift für Berufs- und Wirtschaftspädagogik, 98* (2), 197–224.
- Greimel-Fuhrmann, B. & Geyer, A. (2005). Die Wirkung von Interesse und Sympathie auf die Gesamtbeurteilung in der Lehrevaluation. Direkte und indirekte Effekte unter Berücksichtigung des Lehrerverhaltens. *Empirische Pädagogik, 19* (2), 103–120.

- Greimel-Fuhrmann, B. (2014). Students' Perception of Teaching Behaviour and Its Effect on Evaluation. *International Journal for Cross-Disciplinary Subjects in Education*, 5 (1), 1557–1563. <https://doi.org/10.20533/ijcdse.2042.6364.2014.0218>
- Gruehn, S. (2000). *Unterricht und schulisches Lernen. Schüler als Quellen der Unterrichtsbeschreibung*, Pädagogische Psychologie und Entwicklungspsychologie, Bd. 12. Zugl.: Berlin, Freie Univ., Diss., 1998. Münster: Waxmann.
- Gruehn, S. & Bertram, J. (2014). *Verschiedene Perspektiven auf Unterricht: Wie nehmen Lehrer/innen und Schüler/innen ihren gemeinsamen Unterricht wahr?* Vortrag auf der 2. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Frankfurt am Main.
- Guski, R. (2000). *Wahrnehmung. Eine Einführung in die Psychologie der menschlichen Informationsaufnahme*, Urban-Taschenbücher, Bd. 556, 2., überarb. Aufl.. Stuttgart: Kohlhammer.
- Haag, L. & Lohrmann, K. (2009). Lehrerhandeln: Lehrerkognitionen und Lehrerexpertise. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht*, UTB Schulpädagogik, Pädagogik, Bd. 8423, 2., aktualisierte Aufl., 461–467. Bad Heilbrunn: Klinkhardt.
- Habermas, J. (1988). *Der philosophische Diskurs der Moderne. Zwölf Vorlesungen* (Suhrkamp-Taschenbuch Wissenschaft, Bd. 749, 1. Aufl., [Nachdr.]. Frankfurt am Main: Suhrkamp.
- Habermas, J. (2001). Moralentwicklung und Ich-Identität. In J. Habermas (Hrsg.), *Zur Rekonstruktion des Historischen Materialismus*, Suhrkamp-Taschenbuch Wissenschaft, Bd. 154, 1. Aufl., [Nachdr.], 63ff. Frankfurt am Main: Suhrkamp.
- Habermas, J. (Hrsg.). (2001). *Zur Rekonstruktion des Historischen Materialismus*, Suhrkamp-Taschenbuch Wissenschaft, Bd. 154, 1. Aufl., [Nachdr.]. Frankfurt am Main: Suhrkamp.
- Häcker, H. (2014). Faktorentheorien der Persönlichkeit. In M. A. Wirtz (Hrsg.), *Dorsch. Lexikon der Psychologie*, 17. Aufl., 525. Bern: Hans Huber.
- Haller, M. & Müller, B. (2006). Merkmale der Persönlichkeit und Identität. Ansätze zu ihrer Operationalisierung und Verortung als Erklärungsvariable für Lebenszufriedenheit. *ZUMA Nachrichten*, 30 (59), 9–41. Verfügbar unter <http://www.ssoar.info/ssoar/handle/document/20747>
- Hanfstingl, B. & Mayr, J. (2007). Prognose der Bewährung im Lehrerstudium und im Lehrerberuf. *Journal für Lehrerinnen- und Lehrerbildung*, 7 (2), 48–56.
- Hannover, B. (2015). Geschlecht und soziale Ungleichheit. In H. Reinders (Hrsg.), *Empirische Bildungsforschung. Gegenstandsbereiche*, 2., überarb. Aufl., 201–213. Wiesbaden: Springer VS.
- Harring, J., Stapleton, L. M. & Beretvas, S. N. (Eds.). (2016). *Advances in multilevel modeling for educational research. Addressing practical issues found in real-world applications*

- (CILVR series on latent variable methodology). Charlotte, NC: Information Age Publishing Inc.
- Harris, D. N. & Rutledge, S. A. (2007). *Models and Predictors of Teacher Effectiveness. A review of the literature with lessons from (and for) other occupations*. Unpublished Manuscript, University of Wisconsin at Madison.
- Hart, J. & Driver, J. (1978). Teacher Evaluation as a Function of Student and Instructor Personality. *Teaching of Psychology*, 4 (5), 198–200.
- Harvey, L. & Green, D. (2000). Qualität definieren. Fünf unterschiedliche Ansätze. *Zeitschrift für Pädagogik* (41. Beiheft), 17–39.
- Hattie, J., Beywl, W. & Zierer, K. (2013). *Lernen sichtbar machen*. Baltmannsweiler: Schneider-Verl. Hohengehren.
- Hechinger, M. & Seibert, N. (2015). Eignungsverfahren für angehende (Gymnasial-) Lehrerinnen und Lehrer sind nötig?! In S. Lin-Klitzing, D. Di Fuccia & R. Stengl-Jörns (Hrsg.), *Auf die Lehrperson kommt es an? Beiträge zur Lehrerbildung nach John Hatties "Visible Learning"*, 38-63. Bad Heilbrunn: Klinkhardt.
- Heid, H. (2000). Qualität. Überlegungen zur Begründung einer pädagogischen Beurteilungskategorie. In A. [.] Helmke, W. [.] Hornstein & E. [.] Terhart (Hrsg.), *Qualität und Qualitätssicherung im Bildungsberiech. Überlegungen zur Begründung einer pädagogischen Beurteilungskategorie*, *Zeitschrift für Pädagogik*, 41. Beiheft, 41-51. Weinheim: Beltz.
- Heinrich, M. & Greiner, U. (Hrsg.). (2006). *Schauen, was 'rauskommt. Kompetenzförderung, Evaluation und Systemsteuerung im Bildungswesen*. Wien: Lit Verlag.
- Helmke, A. (1996). Studentische Evaluation der Lehre - Sackgassen und Perspektiven. Anmerkung zum Beitrag von Rosemann & Schweer. *Zeitschrift für Pädagogische Psychologie*, 10 (3/4), 181–186.
- Helmke, A. (2003). Unterrichtsevaluation: Verfahren und Instrumente. *Schulmanagement*, 1, 8–11.
- Helmke, A. (2004). *Unterrichtsqualität. Erfassen, bewerten, verbessern* (3. Aufl.). Seelze: Klett Kallmeyer.
- Helmke, A. (2006). *Unterrichtsqualität erfassen, bewerten, verbessern. Dieses Buch ist Franz-Emanuel Weinert gewidmet* (Schulisches Qualitätsmanagement (6. Aufl.). Seelze: Klett Kallmeyer.
- Helmke, A. (2007). *Was wissen wir über guten Unterricht? Wissenschaftliche Erkenntnisse zur Unterrichtsforschung und Konsequenzen für die Unterrichtsentwicklung*. Landau. Verfügbar unter http://www.bildung.koeln.de/imperia/md/content/selbst_schule/downloads/andreas_helmke_.pdf

- Helmke, A. (2009). Unterrichtsforschung. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht*, UTB Schulpädagogik, Pädagogik, Bd. 8423 (2., aktual. Aufl.), 44–50. Bad Heilbrunn: Klinkhardt.
- Helmke, A. (2014). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts: Franz Emanuel Weinert gewidmet*, Schule weiterentwickeln, Unterricht verbessern Orientierungsband (5. Aufl.). Seelze-Velber: Klett/Kallmeyer.
- Helmke, A. & Jäger, R. S. (Hrsg.). (2002). *Das Projekt MARKUS. Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*. Landau: Verl. Empirische Pädagogik e.V.
- Helmke, A., Schneider, W. & Weinert, F. E. (1986). Quality of instruction and classroom learning outcomes. Results of the German contribution to the Classroom Environment Study of the IEA. *Teaching and Teacher Education*, 2, 1–18.
- Helmke, A. & Schrader, F. W. (1998). Entwicklung im Grundschulalter. Die Münchner Studie "SCHOLASTIK". *Pädagogik*, 50 (6), 25–30.
- Helmke, A. & Schrader, F.-W. (2006). Lehrerprofessionalität und Unterrichtsqualität. Den eigenen Unterricht beurteilen und reflektieren. *Schulmagazin*, 5-10.
- Helmke, A. & Schrader, F.-W. (2008). Merkmale der Unterrichtsqualität: Potenziale, Reichweite und Grenzen. *SEMINAR - Lehrerbildung und Schule*, 3, 17–47.
- Helmke, A. & Schrader, F.-W. (2010). Merkmale der Unterrichtsqualität: Potenzial, Reichweite und Grenzen. In B. Schaal (Hrsg.), *Qualitätssicherung im Bildungswesen. Auftrag und Anspruch der bayerischen Qualitätsagentur; eine Publikation des Staatsinstituts für Schulqualität und Bildungsforschung (ISB)*, 69–110. Münster: Waxmann.
- Helmke, A., Schrader, F.-W., Wagner, W., Nold, G. & Schröder, K. (2008). Selbstkonzept, Motivation und Englischleistung. In E. Klieme (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (Beltz Pädagogik, S. 244–257). Weinheim: Beltz.
- Helmke, A. & Weinert, F. E. (1997). Wissenschaftliche Grundlagen, Realisierungsbedingungen und Ergebnisperspektiven des Längsschnittprojektes SCHOLASTIK. In F. E. Weinert & A. Helmke (Hrsg.), *Entwicklung im Grundschulalter*, 1–12. Weinheim: Beltz Psychologie-Verl.-Union.
- Helmke, A. [., Hornstein, W. [.] & Terhart, E. [.] (Hrsg.). (2000). *Qualität und Qualitätssicherung im Bildungsberiech. Überlegungen zur Begründung einer pädagogischen Beurteilungskategorie*. Zeitschrift für Pädagogik, 41. Beiheft. Weinheim: Beltz.
- Helsper, W. & Böhme, J. (Hrsg.). (2008). *Handbuch der Schulforschung* (2., durchgesehene und erweiterte Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden. <https://doi.org/10.1007/978-3-531-91095-6>

- Henecka, H. P. & Lipowsky, F. (1999). Wege in den Beruf. In H. Ehlert, U. Welbers & V. K. Eckermann (Hrsg.), *Handbuch Praxisinitiativen an Hochschulen. Berufsorientierende Angebote für Studierende an Universitäten*, Hochschulwesen - Wissenschaft und Praxis, 22–23. Neuwied: Luchterhand.
- Hentig, H. v. (1981). *Vom Verkäufer zum Darsteller. Absagen an die Lehrerbildung*, Klett Cotta Sonderdruck: Neue Sammlung, Bd. 2 (2. Aufl.): Klett.
- Hetzer, H. & Arbinger, R. (Hrsg.). (1995). *Angewandte Entwicklungspsychologie des Kindes- und Jugendalters*, UTB für Wissenschaft, Bd. 935 (3. Aufl., unveränd. Nachdr. der 2., überarb. Aufl.). Heidelberg: Quelle & Meyer.
- Hill, H. C., Rowan, B. & Ball, D. L. (2005). Effects of Teachers' Mathematical Knowledge for Teaching on Student Achievement. *American Educational Research Journal*, 42 (2), 371–406. <https://doi.org/10.3102/00028312042002371>
- Hofer, M. (Hrsg.). (1981). *Informationsverarbeitung und Entscheidungsverhalten von Lehrern* (U&S Pädagogik). München: Urban & Schwarzenberg.
- Hoffmann, L. (Ed.). (1998). *Interest and learning. Proceedings of the Seeon Conference on Interest and Gender* (IPN, vol. 164). Kiel: IPN.
- Hofmann, W., Gschwendner, T. & Schmitt, M. (2009). The road to the unconscious self not taken. Discrepancies between self- and observer-inferences about implicit dispositions from nonverbal behavioural cues. *European Journal of Personality*, 23 (4), 343–366. <https://doi.org/10.1002/per.722>
- Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality*, 8 (3), 149–162. <https://doi.org/10.1002/per.2410080302>
- Hogan, R. & Roberts, B. W. (2004). A Socioanalytic Model of Maturity. *Journal of Career Assessment*, 12 (2), 207–217. <https://doi.org/10.1177/1069072703255882>
- Hogan, R. T. (1991). Personality and personality measurement. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (2. ed.), 873–919. Palo Alto/Calif.: Consulting Psychologists Press. Verfügbar unter <http://psycnet.apa.org/psycinfo/1993-97200-013>
- Hohensinn, C., Kubinger, K. D., Reif, M., Schleicher, E. & Khorramdel, L. (2011). Analysing item position effects due to test booklet design within large-scale assessment. *Educational Research and Evaluation*, 17 (6), 497–509. <https://doi.org/10.1080/13803611.2011.632668>
- Holland, J. L. (1997). *Making vocational choices. A theory of vocational personalities and work environments* (3. ed.). Lutz, Fla.: Psychological Assessment Resources.
- Holling, H. (Hrsg.). (2010). *Handbuch Statistik, Methoden und Evaluation*, Handbuch der Psychologie, Bd. 13. Göttingen: Hogrefe.

- Holtappels, H. G. & Höhmann, K. (Hrsg.). (2005). *Schulentwicklung und Schulwirksamkeit. Systemsteuerung, Bildungschancen und Entwicklung der Schule*. Weinheim: Juventa.
- Holtappels, H. G., Klieme, E., Rauschenbach, T. & Stecher, L. (Hrsg.). (2007). *Ganztags-schule in Deutschland. Ergebnisse der Ausgangserhebung der "Studie zur Entwicklung von Ganztags-schulen" (StEG)*. Weinheim: Juventa.
- Holtz, B. C., Ployhart, R. E. & Dominguez, A. (2005). Testing the Rules of Justice. The Effects of Frame-of-Reference and Pre-Test Validity Information on Personality Test Responses and Test Perceptions. *International Journal of Selection and Assessment*, 13 (1), 75–86. <https://doi.org/10.1111/j.0965-075X.2005.00301.x>
- Homburg, C. & Baumgartner, H. (1995). Beurteilung von Kausalmodellen. Bestandsauf-nahme und Anwendungsempfehlungen. *Zeitschrift für Forschung und Praxis*, 162–176.
- Horsch, L. (2015). *Reliabilität und Validität des OPD-Strukturfragebogens (OPD-SF)*. Kurzfassung einer medizinischen Dissertation, Ruprecht-Karls-Universität Heidelberg. Heidelberg.
- Hosenfeld, I., Helmke, A., Ridder, A. & Schrader, F. W. (2001). Eine mehrbenenanalytische Betrachtung von Schul- und Klasseneffekten. *Empirische Pädagogik*, 15 (4), 513–534.
- Hossiep (2014). Fünf-Faktoren-Modell. In M. A. Wirtz (Hrsg.), *Dorsch. Lexikon der Psycho-logie* (17. Aufl.). Bern: Hans Huber.
- Hossiep, R. & Paschen, M. (2003). *BIP - Bochumer Inventar zur berufsbezogenen Persönlich-keitsbestimmung* (2., vollständig überarbeitete Auflage). Göttingen: Hogrefe. Verfügbar unter <https://www.hogrefe.de/shop/bochumer-inventar-zur-berufsbezogenen-persoendlich-keitsbeschreibung.html>
- Höttecke, D. (Hrsg.). (2007). *Naturwissenschaftlicher Unterricht im internationalen Ver-gleich. [f tit und Physik, Bd. 27; 33. Jahrestagung der GDGP vom 18. bis 21. September 2006 in den Räumen der Universität Bern]*. Berlin: LIT-Verl.
- Hox, J. J. (1995). *Applied multilevel analysis*, Electronic ed., [Elektronische Ressource]. Am-sterdam: TT-Publikaties.
- Hox, J. J. (2010). *Multilevel analysis. Techniques and applications* (Quantitative methodology series, 2. ed.). New York, NY: Routledge.
- Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis. Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidis-ciplinary Journal*, 6 (1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hugener, I., Pauli, C. & Reusser, K. (2006). Videoanalyse. Teil 3. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur Schwei-zerisch-Deutschen Videostudie. Unterrichtsqualität, Lernverhalten und mathematisches Verständnis*. Frankfurt am Main: GFPPF.

- Hülsheger, U., Specht, E. & Spinath, F. (2006). Validität des BIP und des NEO-PI-R. Wie geeignet sind ein berufsbezogener und ein nicht explizit berufsbezogener Persönlichkeits-test zur Erklärung von Berufserfolg? *Zeitschrift für Arbeits- und Organisationspsychologie*, 50 (3), 135–147.
- Hunthausen, J. M., Truxillo, D. M., Bauer, T. N. & Hammer, L. B. (2003). A field study of frame-of-reference effects on personality test validity. *Journal of Applied Psychology*, 88 (3), 545–551. <https://doi.org/10.1037/0021-9010.88.3.545>
- Hurtz, G. M. & Donovan, J. J. (2000). Personality and job performance. The Big Five revisited. *Journal of Applied Psychology*, 85 (6), 869–879. <https://doi.org/10.1037//0021-9010.85.6.869>
- Husén, T. & Postlethwaite, T. (Eds.). (1993). *The international encyclopedia of education*. Oxford: Pergamon.
- Iser, J. & Schmidt, P. (2005). Werte und Big Five. Trennbarkeit der Konzepte und Erklärungskraft für politische Orientierungen. In S. Schumann (Hrsg.), *Persönlichkeit. Eine vergessene Größe der empirischen Sozialforschung* (1. Aufl.), 279–299. Wiesbaden: VS Verl. für Sozialwiss.
- Jackson, D. N. (1974). *Personality research form manual*. Port Huron, Mi.: Research Psychologists Press.
- Janich, N. & Thim-Mabrey, C. (Hrsg.). (2003). *Sprachidentität - Identität durch Sprache*, Tübinger Beiträge zur Linguistik, Bd. 465. Tübingen: Narr.
- Jansen, M., Jensen, P. E. & Mylow, P. (1972). Teacher characteristics and other factors affecting classroom interaction and teaching behaviour. *International Review of Education*, 18 (1), 529–540.
- Jolijn Hendriks, A. A., Perugini, M., Angleitner, A., Ostendorf, F., Johnson, J. A., Fruyt, F. de et al. (2003). The five-factor personality inventory. Cross-cultural generalizability across 13 countries. *European Journal of Personality*, 17 (5), 347–373. <https://doi.org/10.1002/per.491>
- Jones, E. E. (Ed.). (1972). *Attribution. Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.
- Jones, E. E., Kanouse, D. E. & Kelley, H. H. (Eds.). (1987). *Attribution. Perceiving the causes of behavior* (UMI Out-of-print books on demand). Ann Arbor, Hillsdale, N.J.: Bell & Howell Erlbaum.
- Jones, E. E. & Nisbett, R. E. (1972). The actor and the observer. Divergent perceptions of the causes of behavior. In E. E. Jones (Ed.), *Attribution. Perceiving the causes of behaviour*, 79–94. Morristown, NJ: General Learning Press.

- Jones, E. E. & Nisbett, R. E. (1987). The Actor and the Observer. Divergent perceptions of the causes of behaviour. In E. E. Jones, D. E. Kanouse & H. H. Kelley (Eds.), *Attribution. Perceiving the causes of behavior* (UMI Out-of-print books on demand), 79–94. Ann Arbor, Hillsdale, N.J.: Bell & Howell Erlbaum.
- Jones, J., Lennie, M. & Robinson, A. (1985). *Students' expectations of good teachers: Primary, secondary and tertiary views*. Paper presented at NZARE Conference, Auckland, New Zealand.
- Jones, J. (1989). Students' ratings of teacher personality and teaching competence. *Higher Education, 18* (5), 551–558. <https://doi.org/10.1007/BF00138747>
- Jordan, A., Ross, N., Krauss, S., Baumert, J., Blum, W., Neubrand, M. et al. (2006). *Klassifikationsschema für Mathematikaufgaben. Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt* (Materialien aus der Bildungsforschung, Nr. 81). Berlin: Max-Planck-Inst. für Bildungsforschung.
- Joreskog, K. G. & Sorbom, D. (1982). Recent Developments in Structural Equation Modeling. *Journal of Marketing Research, 19* (4), 404. <https://doi.org/10.2307/3151714>
- Judge, T. A., Heller, D. & Mount, M. K. (2002). Five-factor model of personality and job satisfaction. A meta-analysis. *Journal of Applied Psychology, 87* (3), 530–541. <https://doi.org/10.1037/0021-9010.87.3.530>
- Jung, C. G. & Baynes, H. G. (1924). PSYCHOLOGICAL TYPES: Translated with an Introduction. *Journal of Nervous and Mental Disease, 60* (2), 221–224.
- Kastens, C. (2009). *Handlungsmotivation und Unterrichtsqualität: Erfassung affektiv-motivationaler Prädiktoren des Unterrichtshandelns von Lehrkräften*. Dissertation, Christian-Albrechts-Universität zu Kiel. Kiel.
- Keaveny, T. J. & McGann, A. F. (1978). Behavioral dimensions associated with students' global ratings of college professors. *Research in Higher Education, 9* (4), 333–345. <https://doi.org/10.1007/BF00991405>
- Keller, M. M., Götz, T., Becker, E. S. & Morger, V. (2013). *Lehrerenthusiasmus und Expressivität: Zwei Seiten einer Medaille?* Vortrag auf der 1. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Kiel.
- Kenny, D. A. & DePaulo, B. M. (1993). Do people know how others view them? An empirical and theoretical account. *Psychological Bulletin, 114* (1), 145–161. <https://doi.org/10.1037/0033-2909.114.1.145>
- Kerschensteiner, G. (1926). *Theorie der Bildung*. Wiesbaden: Vieweg+Teubner Verlag. <https://doi.org/10.1007/978-3-663-15795-3>
- Kersting, M. (2005). Beratung und Auswahl von Studienbewerbern. Ziele und Methoden. *Psychologische Rundschau, 56* (2), 149–150. <https://doi.org/10.1026/0033-3042.56.2.149>

- Kessels, U. (2005). Fitting into the stereotype. How gender-stereotyped perceptions of prototypic peers relate to liking for school subjects. *European Journal of Psychology of Education*, 20 (3), 309–323. <https://doi.org/10.1007/BF03173559>
- Keupp, H. & Hohl, J. (Hrsg.). (2006). *Subjektdiskurse im gesellschaftlichen Wandel. Zur Theorie des Subjekts in der Spätmoderne (Sozialtheorie)*. Bielefeld: Transcript-Verlag. <https://doi.org/10.14361/9783839405628>
- Kleickmann, T. & Möller, K. (2007). Haben Lehrerfortbildungen einen Effekt auf die Lernzuwächse bei Schülerinnen und Schülern? In D. Höttecke (Hrsg.), *Naturwissenschaftlicher Unterricht im internationalen Vergleich. [f tit und Physik, Bd. 27; 33. Jahrestagung der GDGP vom 18. bis 21. September 2006 in den Räumen der Universität Bern]*, 506–508. Berlin: LIT-Verl.
- Klieme, E. (2002). Was ist guter Unterricht? Ergebnisse der TIMSS-Videostudie im Fach Mathematik. In W. Bergsdorf (Hrsg.), *Herausforderungen der Bildungsgesellschaft*, 89–113. Weimar: Rhino-Verlag.
- Klieme, E. (2006). Empirische Unterrichtsforschung: Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 52 (6), 765–773.
- Klieme, E. (Hrsg.). (2008). *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie (Beltz Pädagogik)*. Weinheim: Beltz.
- Klieme, E. (2013). Qualitätsbeurteilung von Schule und Unterricht. Möglichkeiten und Grenzen einer begriffsanalytischen Reflexion. *Zeitschrift für Erziehungswissenschaft*, 2 (16), 433–441. Ein Kommentar zu Helmut Heid.
- Klieme, E., Lipowsky, F., Rakoczy, K. & Ratzka, N. (2006). Qualitätsdimensionen und Wirksamkeit von Mathematikunterricht. Theoretische Grundlagen und ausgewählte Ergebnisse des Projekts "Pythagoras". In M. Prenzel & L. Allolio-Näcke (Hrsg.), *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms*, 127–146. Münster: Waxmann.
- Klieme, E., Pauli, C. & Reusser, K. (Hrsg.). (2005). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur Schweizerisch- Deutschen Videostudie. Unterrichtsqualität, Lernverhalten und mathematisches Verständnis*, Materialien zur Bildungsforschung, Bd. 13. Frankfurt am Main: GFPPF.
- Klieme, E., Pauli, C. & Reusser, K. (Hrsg.). (2006). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur Schweizerisch- Deutschen Videostudie. Unterrichtsqualität, Lernverhalten und mathematisches Verständnis*. Frankfurt am Main: GFPPF.

- Klieme, E. & Rakoczy, K. (2008). Empirische Unterrichtsforschung und Fachdidaktik. Outcome-orientierte Messung und Prozessqualität des Unterrichts. *Zeitschrift für Pädagogik*, 54 (2), 222–237.
- Klieme, E. & Reusser, K. (2003). Unterrichtsqualität und mathematisches Verständnis im internationalen Vergleich. Ein Forschungsprojekt und erste Schritte zur Realisierung. *Unterrichtswissenschaft*, 31 (3), 194–205.
- Klieme, E. & Tippelt, R. (Hrsg.). (2008). *Qualitätssicherung im Bildungswesen. Eine aktuelle Zwischenbilanz*, Zeitschrift für Pädagogik. Beiheft, Bd. 53. Weinheim: Beltz.
- Klusmann, U., Kunter, M., Trautwein, U. & Baumert, J. (2006). Lehrerbelastung und Unterrichtsqualität aus der Perspektive von Lehrenden und Lernenden. *Zeitschrift für Pädagogische Psychologie*, 20 (3), 161–173. <https://doi.org/10.1024/1010-0652.20.3.161>
- Kneipp, L. B., Kelly, K. E., Biscoe, J. D. & Richard, B. (2010). The impact of instructor's personality characteristics on quality of instruction. *College Student Journal*, 44 (4), 901–906.
- Köller, O., Watermann, R., Trautwein, U. & Lüdtke, O. (Hrsg.). (2004). *Wege zur Hochschulreife in Baden-Württemberg. TOSCA - Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien*. Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-322-80906-3>
- Kounin, J. S. (1976). *Techniken der Klassenführung*. Stuttgart: Klett.
- Kounin, J. S. (2006). *Techniken der Klassenführung*, Standardwerke aus Psychologie und Pädagogik Reprints, Bd. 3. Münster: Waxmann.
- Krammer, K., Ratzka, N., Klieme, E., Lipowsky, F., Pauli, C. & Reusser, K. (2006). Learning with classroom videos. Conception and first results of an online teacher-training program. *ZDM*, 38 (5), 422–432. <https://doi.org/10.1007/BF02652803>
- Krapp, A. (1992). Interesse, Lernen und Leistung. Neue Forschungsansätze in der Pädagogischen Psychologie. *Zeitschrift für Pädagogik*, 38 (5), 747–770.
- Krapp, A. (1999). Intrinsische Lernmotivation und Interesse. Forschungsansätze und konzeptuelle Überlegungen. *Zeitschrift für Pädagogik*, 45 (3), 387–406.
- Krapp, A. (2014). Interessenskonstrukt. In M. A. Wirtz (Hrsg.), *Dorsch. Lexikon der Psychologie* (17. Aufl.). Bern: Hans Huber.
- Krenz, A. (2010). Die Persönlichkeit der Erzieherin. Dreh- und Angelpunkt jeder "guten" Pädagogik. In A. Krenz (Hrsg.), *Kindorientierte Elementarpädagogik*, Frühe Bildung und Erziehung (1. Aufl.), 198–206. Göttingen: Vandenhoeck & Ruprecht.
- Krenz, A. (Hrsg.). (2010). *Kindorientierte Elementarpädagogik*, Frühe Bildung und Erziehung (1. Aufl.). Göttingen: Vandenhoeck & Ruprecht.
- Kröhne, U. & Frey, A. (Hrsg.). (2013). *Psychological Test and Assessment Modeling*. Volume 55, Issue 1: Pabst Science Publishers.

- Kuncel, N. R. & Hezlett, S. A. (2010). Fact and Fiction in Cognitive Ability Testing for Admissions and Hiring Decisions. *Current Directions in Psychological Science*, 19 (6), 339–345. <https://doi.org/10.1177/0963721410389459>
- Künsting, J., Billich, M. & Lipowsky, F. (2009). Der Einfluss von Lehrerkompetenzen und Lehrerhandeln auf dem Schulerfolg von Lernenden. In O. Zlatkin-Troitschanskaia (Hrsg.), *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung*, Beltz-Bibliothek, 655–667. Weinheim: Beltz.
- Kunter, M. (2005). *Multiple Ziele im Mathematikunterricht*, Pädagogische Psychologie und Entwicklungspsychologie, Bd. 51. Zugl.: Berlin, Freie Univ., Diss., 2004. Münster: Waxmann.
- Kunter, M. & Baumert, J. (2010). Einführung in den Themenschwerpunkt "Lehrerforschung". *Unterrichtswissenschaft*, 38 (1), 3–4.
- Kunter, M., Klusmann, U., Baumert, J., Richter, D., Voss, T. & Hachfeld, A. (2013). Professional competence of teachers. Effects on instructional quality and student development. *Journal of Educational Psychology*, 105 (3), 805–820. <https://doi.org/10.1037/a0032583>
- Labudde, P. (2009). *Fachunterricht und fächerübergreifender Unterricht: Grundlagen*. In K.-H. Arnold, U. Sandfuchs & J. Wiechmann (Hrsg.), *Handbuch Unterricht*, UTB Schulpädagogik, Pädagogik, Bd. 8423, (2., aktualisierte Aufl.), 331-336. Bad Heilbrunn: Klinkhardt.
- Lanahan, L., McGrath, D. J., McLaughlin, M., Burian-Fitzgerald, M. & Salganik, L. (2005). *Fundamental problems in the measurement of instructional processes: Estimating reasonable effect sizes and conceptualizing what is important to measure*. Washington, DC.: American Institutes for Research.
- Lang, F. R. & Lüdtke, O. (2005). *Der Big Five-Ansatz der Persönlichkeitsforschung: Instrumente und Vorgehen*.
- Lang, F. R., Lüdtke, O. & Asendorpf, J. B. (2001). Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen. *Diagnostica*, 47 (3), 111–121. <https://doi.org/10.1026//0012-1924.47.3.111>
- Lehrke, M. & Hofmann, L. (Hrsg.). (1987). *Schülerinteressen am naturwissenschaftlichen Unterricht. Untersuchungen und Erklärungen; Beiträge zum 12. IPN-Symposion; die vorl. Arbeit entstand am Inst. für die Pädagogik der Naturwissenschaften (IPN) an der Univ. Kiel*, Didaktik der Naturwissenschaften, Bd. 12. Köln: Aulis-Verl. Deubner.
- Leschinsky, A. (Hrsg.). (1996). *Die Institutionalisierung von Lehren und Lernen. Beiträge zu einer Theorie der Schule; Festgabe für Peter Martin Roeder*. Weinheim: Beltz.
- L'Hommedieu, R., Menges, R. J. & Brinko, K. T. (1990). Methodological explanations for the modest effects of feedback from student ratings. *Journal of Educational Psychology*, 82 (2), 232–241. <https://doi.org/10.1037/0022-0663.82.2.232>

- Lievens, F., Corte, W. de & Schollaert, E. (2008). A Closer Look at the Frame-of-Reference Effect in Personality Scale Scores and Validity. *Journal of Applied Psychology*, 93 (2), 268–279.
- Link, J. (2006). *Versuch über den Normalismus. Wie Normalität produziert wird.* (3. Aufl.). Göttingen: Vandenhoeck & Ruprecht. <https://doi.org/10.1515/zrs.2009.017>
- Lipowsky, F. (1999). Lernzeit und Konzentration. Grundschul Kinder in offenen Lernsituationen. *Die Deutsche Schule*, 91 (2), 232–245.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Empirische Evidenzen für Zusammenhänge zwischen Lehrerkompetenzen, Lehrerhandeln und dem Lernen der Schüler. In C. Allemann-Ghionda & E. Terhart (Hrsg.), *Kompetenzen und Kompetenzentwicklung von Lehrerinnen und Lehrern. Ausbildung und Beruf*, Zeitschrift für Pädagogik. Beihefte, Bd. 51, 47–70. Weinheim: Beltz.
- Lipowsky, F. (2006). Auf den Lehrer kommt es an. Exzerpt. *Zeitschrift für Pädagogik*, 52, 47–70.
- Lipowsky, F. (2007). Unterrichtsqualität in der Grundschule. Ansätze und Befunde der nationalen und internationalen Forschung. In K. Möller, P. Hanke, C. Beinbrech, T. Kleickmann & R. Schages (Hrsg.), *Qualität von Grundschulunterricht entwickeln, erfassen und bewerten* (1. Aufl.), 35–49. Wiesbaden: Springer VS.
- Lipowsky, F. (2009). Unterricht. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie*, Springer-Lehrbuch, 73–102 Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg.
- Lipowsky, F. (2010). Lernen im Beruf. Empirische Befunde zur Wirksamkeit von Lehrerfortbildungen. In F. H. Müller, A. Eichenberger, M. Lüders & J. Mayr (Hrsg.), *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung*, 51–72. Münster: Waxmann.
- Lipowsky, F., Drolling-Vetter, B., Hartig, J. & Klieme, E. (2006). Leistungstests. Teil 2. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur Schweizerisch- Deutschen Videostudie. Unterrichtsqualität, Lernverhalten und mathematisches Verständnis*. Frankfurt am Main: GFPPF.
- Lipowsky, F., Rakoczy, K., Klieme, E., Reusser, K. & Pauli, C. (2005). Unterrichtsqualität im Schnittpunkt unterschiedlicher Perspektiven. Rahmenkonzept und erste Ergebnisse einer binationalen Studie zum Mathematikunterricht in der Sekundarstufe I. In H. G. Holtappels & K. Höhmann (Hrsg.), *Schulentwicklung und Schulwirksamkeit. Systemsteuerung, Bildungschancen und Entwicklung der Schule*, 223–238. Weinheim: Juventa.
- Lublin, J. & Barrand, J. (1987). What the learner can teach us. *Research and Development in Higher Education*, 9, 5–11.

- Lüdtke, O., Robitzsch, A., Trautwein, U. & Köller, O. (2007). Umgang mit fehlenden Werten in der psychologischen Forschung. *Psychologische Rundschau*, 58 (2), 103–117. <https://doi.org/10.1026/0033-3042.58.2.103>
- Lunk, G. (1927). *Das Interesse. Philosophisch-pädagogischer Teil* (Bd. 2). Bad Heilbrunn: Klinkhardt.
- Maaß, S. (2014). Normalisierte Körper und hypernormale Subjekte. Technologien des Körpers und des Selbst in der Castingshow Germany's next Topmodel. *Zeitschrift für angewandte Diskurstheorie*, 66/67 (1/2), 101–114.
- Magno, C. & Sembrano, J. (2008). The Role of Teacher Efficacy and Characteristics on Teaching Effectiveness, Performance and Use of Learner-centered Practices. *The Asia Pacific Education Researcher*, 16, 73–90. Verfügbar unter https://pdfs.semanticscholar.org/560f/655b56941300c8c63066a76c40f1318b387d.pdf?_ga=2.30045948.277986174.1553699878-1373830535.1553699878
- Malle, B. F. (2006). The actor-observer asymmetry in attribution: a (surprising) meta-analysis. *Psychological Bulletin*, 132 (6), 895–919. <https://doi.org/10.1037/0033-2909.132.6.895>
- Marques, T. E., Lane, D. M. & Dorfman, P. W. (1979). Toward the development of a system for instructional evaluation. Is there consensus regarding what constitutes effective teaching? *Journal of Educational Psychology*, 71 (6), 840–849. <https://doi.org/10.1037/0022-0663.71.6.840>
- Marsh, H. W. & Dunkin, M. J. (1992). Students' Evaluations of University Teaching: A Multidisciplinary Perspective. In J. C. Smart (Hrsg.), *Higher Education: Handbook of Theory and Research* (8. Aufl.), 143–233. Netherlands: Springer.
- Marsh, H. W. & Roche, L. (1993). The Use of Students' Evaluations and an Individually Structured Intervention to Enhance University Teaching Effectiveness. *American Educational Research Journal*, 30 (1), 217–251. <https://doi.org/10.3102/00028312030001217>
- Marsh, H. W. (1984). Students' evaluations of university teaching. Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology*, 76 (5), 707–754. <https://doi.org/10.1037/0022-0663.76.5.707>
- Marsh, H. W. & Hocevar, D. (1991). Students' evaluations of teaching effectiveness. The stability of mean ratings of the same teachers over a 13-year period. *Teaching and Teacher Education*, 7 (4), 303–314. [https://doi.org/10.1016/0742-051X\(91\)90001-6](https://doi.org/10.1016/0742-051X(91)90001-6)
- Mayr, J. (1994). Junge LehrerInnen: Interessen, Kompetenzen und Befindlichkeit - und zwei Blicke zurück. In J. Mayr (Hrsg.), *Lehrer/in werden*, 177–199. Innsbruck: Österreichischer Studienverlag.
- Mayr, J. (Hrsg.). (1994). *Lehrer/in werden*. Innsbruck: Österreichischer Studienverlag.

- Mayr, J. (2000). Schriftliche Informationen, Selbsterkundungsverfahren und Tests als Hilfsmittel der Laufbahnberatung. In B. Sieland & B. Rißland (Hrsg.), *Qualitätssicherung in der Lehrerbildung. Lehrerarbeit: Bedingungsfaktoren und Qualitätskriterien*, Studien zur Erwachsenenbildung, Bd. 9, 233–265. Hamburg: Kovač.
- Mayr, J. (2003). Persönlichkeitsfragebögen in der Lehrerforschung und Lehrerberatung. *Theorie & Praxis Texte zur Lehrerbildung* (20), 81.
- Mayr, J. (2009). LehrerIn werden in Österreich: empirische Befunde zum Lehramtsstudium. *Erziehung & Unterricht*, 159, 14–33.
- Mayr, J. (2014). Der Persönlichkeitsansatz in der Lehrerforschung. In E. Terhart, H. Bennewitz & M. Rothland (Hrsg.), *Handbuch der Forschung zum Lehrerberuf*, 125–148. Münster: Waxmann.
- Mayr, J. (2019). *Forschungsnotiz vom 10.02.2019*. (Unveröffentlichtes Manuskript).
- Mayr, J. & Neuweg, G. H. (2006). Der Persönlichkeitseinsatz in der Lehrer/innen/forschung. Grundsätzliche Überlegungen, exemplarische Befunde und Implikationen für die Lehrer/innen/bildung. In M. Heinrich & U. Greiner (Hrsg.), *Schauen, was 'rauskommt. Kompetenzförderung, Evaluation und Systemsteuerung im Bildungswesen*, 183–206. Wien: Lit Verlag.
- Mayr, J., & Neuweg, G. H. (2018). Die unterrichtsmethodische Grundeinstellung kaufmännischer Lehrpersonen im Spannungsfeld von Instruktivismus und Konstruktivismus. In B. Greimel-Fuhrmann (Hrsg.), *Wirtschaftspädagogische Forschung und Impulse für die Wirtschaftsdidaktik. Beiträge zum 12. Österreichischen Wirtschaftspädagogikkongress am 26.4.2018 in Wien*. Verfügbar unter www.bwpat.de/wipaed-at1/neuweg_mayr_wipaed-at_2018.pdf
- McCrae, R. R. & Costa, P. T. (1987). Validation of the Five-Factor Model of Personality Across Instruments and Observers. *Journal of Personality and Social Psychology*, 52 (1), 81–90.
- McCrae, R. R. & John, O. P. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality*, 60 (2), 175–215. <https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>
- McCrae, R. R. & Weiss, A. (2007). Observer Ratings of Personality. In R. W. Robins, R. C. Fraley & R. F. Krueger (Hrsg.), *Handbook of Research Methods in Personality Psychology*, 259–272. New York: The Guilford Press.
- McKeachie, W. J. (1979). Student Ratings of Faculty. A Reprise. *Academe*, 65 (6), 384. <https://doi.org/10.2307/40248725>
- Mead, G. H. (1934). *Mind, Self and Society*. Chicago: University of Chicago Press.

- Mecheril, P. (2006). Das unmögliche Subjekt. In H. Keupp & J. Hohl (Hrsg.), *Subjektdiskurse im gesellschaftlichen Wandel. Zur Theorie des Subjekts in der Spätmoderne*, Sozialtheorie, 119–141. Bielefeld: Transcript-Verlag.
- Meyer, H. (2004). *Was ist guter Unterricht?* Berlin: Cornelsen.
- Meyer, H. (2016). *Was ist guter Unterricht?* (11. Auflage). Berlin: Cornelsen.
- Meyer-Drawe, K. (Hrsg.). (2000). *Illusionen von Autonomie. Diesseits von Ohnmacht und Allmacht des Ich* (2. Aufl.). München: Kirchheim.
- Millman, J. (Ed.). (1981). *Handbook of teacher evaluation*. Beverly Hills: Sage.
- Mintzes, J. J. (1979). Overt Teaching Behaviors and Student Ratings of Instructors. *The Journal of Experimental Education*, 48 (2), 145–153. <https://doi.org/10.1080/00220973.1979.11011728>
- Mischel, W. (2004). Toward an integrative science of the person. *Annual review of psychology*, 55, 1–22. <https://doi.org/10.1146/annurev.psych.55.042902.130709>
- Möller, K., Hanke, P., Beinbrech, C., Kleickmann, T. & Schages, R. (Hrsg.). (2007). *Qualität von Grundschulunterricht entwickeln, erfassen und bewerten* (1. Aufl.). Wiesbaden: Springer VS.
- Moos, R. H. & Trickett, E. J. (1974). *Classroom Environment Scale Manual*. Palo Alto/Calif.: Consulting Psychologists Press.
- Morgenstern, C. & Keeves, J. P. (1993). Descriptive Scales for Measuring Educational Climate. In T. Husén & T. Postlethwaite (Eds.), *The international encyclopedia of education* (Bd. 3), 1475–1483. Oxford: Pergamon.
- Morris, C. W., Huebner, D. R., Joas, H. & Mead, G. H. (Eds.). (2015). *Mind, self, and society. The definitive edition* (Annotated edition). Chicago: The University of Chicago Press.
- Moser, K. & Soucek, R. (2003). Das Bochumer Inventar zur berufsbezogenen Persönlichkeitsbeschreibung (BIP). *Zeitschrift für Personalpsychologie*, 2, 101–104.
- Mroczek, D. K. & Little, T. D. (Eds.). (2006). *Handbook of personality development*. Mahwah, NJ: Erlbaum.
- Müller, F. H., Eichenberger, A., Lüders, M. & Mayr, J. (Hrsg.). (2010). *Lehrerinnen und Lehrer lernen. Konzepte und Befunde zur Lehrerfortbildung*. Münster: Waxmann.
- Murray, F. B. (Hrsg.). (1996). *The teacher educator's handbook. Building a knowledge base for the preparation of teachers* (The Jossey-Bass education series, 1. ed.). San Francisco: Jossey-Bass.
- Murray, H. G. (1975). Predicting Student Ratings of College Teaching from Peer Ratings of Personality Types. *Teaching of Psychology*, 2 (2), 66–69. Verfügbar unter http://www.tandfonline.com/doi/abs/10.1207/s15328023top0202_4?journalCode=htop20

- Murray, H. G. (1978). *Teacher Ratings, Student Achievement, and Teacher Personality Traits*. Annual Meeting of the Canadian Psychological Association.
- Murray, H. G. (1983). Low-inference classroom teaching behaviors and student ratings of college teaching effectiveness. *Journal of Educational Psychology*, 75 (1), 138–149. <https://doi.org/10.1037/0022-0663.75.1.138>
- Murray, H. G. & Lawrence, C. (1980). Speech and drama training for lectures as a means of improving university teaching. *Research in Higher Education*, 13 (1), 73–90. <https://doi.org/10.1007/BF00975777>
- Murray, H. G., Rushton, J. P. & Paunonen, S. V. (1990). Teacher personality traits and student instructional ratings in six types of university courses. *Journal of Educational Psychology*, 82 (2), 250–261. <https://doi.org/10.1037/0022-0663.82.2.250>
- Muthén, L. K. & Muthén, B. O. (2012). *Mplus. Statistical Analysis with Latent Variables - User's Guide* (Seventh Edition). Los Angeles, CA: Muthén & Muthén.
- Muthén, L. K. & Muthén, B. O. (2012) Mplus statistical modeling software [Computer software]. Los Angeles, CA.
- Naftulin, D. H., Ware Jr, J. E. & Donnelly, F. A. ... (1973). The Doctor Fox Lecture: a paradigm of educational seduction. *Journal of Medical Education*, 48 (7), 630–635.
- Neubauer, A., Koschmieder, C., Krammer, G., Mayr, J., Müller, F. H., Pflanzl, B. et al. (2017). TESAT – Ein neues Verfahren zur Eignungsfeststellung und Bewerberauswahl für das Lehramtsstudium. *Zeitschrift für Bildungsforschung*, 9 (1–2), 9. <https://doi.org/10.1007/s35834-017-0178-3>
- Neuber, V., Künsting, J. & Lipowsky, F. (2014). *Beeinflusst der Enthusiasmus einer Lehrperson deren unterrichtliches Handeln?* DGfE-Kongress, Berlin.
- Neuweg, G. H. & Mayr, J. (2018). Die unterrichtsmethodische Grundeinstellung kaufmännischer Lehrpersonen im Spannungsfeld von Instruktivismus und Konstruktivismus. In B. Greimel-Fuhrmann (Hrsg.), *Wirtschaftspädagogische Forschung und Impulse für die Wirtschaftsdidaktik. Beiträge zum 12. Österreichischen Wirtschaftspädagogikkongress am 26.4.2018 in Wien*. Verfügbar unter www.bwpat.de/wipaed-at1/neuweg_mayr_wipaed-at_2018.pdf
- Neyer, F. (2014). Fremdbbericht. In M. A. Wirtz (Hrsg.), *Dorsch. Lexikon der Psychologie* (17. Aufl.). Bern: Hans Huber.
- Nieskens, B. (2002). "Einblicke" als Entwicklungsanstoß. *Journal für Lehrerinnen- und Lehrerbildung*, 2 (2), 56–61.
- Nonaka, I. & Takeuchi, H. (1997). *Die Organisation des Wissens. Wie japanische Unternehmen eine brachliegende Ressource nutzbar machen*. Frankfurt/Main: Campus-Verl.

- Norzila, A. R., Parilah, M. S. & Fauziah, A. (2007). Perceived and preferred teaching styles (methods) of English for specific purposes (ESP) students. *Social Science and Humanities Research*, 2 (2), 1–20.
- Oerter, R. & Montada, L. (Hrsg.). (2002). *Entwicklungspsychologie. [Lehrbuch]* (5., vollst. überarb. Aufl.). Weinheim: Beltz PVU.
- Oppenrieder, W. & Thurmair, M. (2003). Sprachidentität im Kontext von Mehrsprachigkeit. In N. Janich & C. Thim-Mabrey (Hrsg.), *Sprachidentität - Identität durch Sprache*, Tübinger Beiträge zur Linguistik, Bd. 465, 39–60. Tübingen: Narr.
- Ostendorf, F. & Angleitner, A. (2004). *NEO-Persönlichkeitsinventar (revidierte Form, NEO-PI-R). Nach Paul T. Costa und Robert R. McCrae*. Göttingen: Hogrefe.
- Othman, F. B. (2009). *A Study on Personality that Influences Teaching Effectiveness*. Masterarbeit.
- Paetow, B.-P. (2004). *Nicht-Identität als Bezugspunkt von Bildungsprozessen. Eine interkulturelle Studie zum (Mahayana-)Buddhismus aus erziehungswissenschaftlicher Sicht*. Dissertation, Universität Bielefeld.
- Paulhus, D. L. & Vazire, S. (2007). The self-report method. In R. W. Robins, R. C. Fraley & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology*, 224–239. New York: Guilford Press.
- Pauli, C. & Reusser, K. (2006). Von international vergleichenden Video Surveys zur videobasierten Unterrichtsforschung und -entwicklung. *Zeitschrift für Pädagogik*, 52 (6), 774–798.
- Pelikan, H., Demmer, W. & Hurrelmann, K. (Hrsg.). (1993). *Gesundheitsförderung in Organisationen*. Weinheim: Juventa.
- Pervin, L. A. (2000). *Persönlichkeitstheorien. Freud, Adler, Jung, Rogers, Kelly, Cattell, Eysenck, Skinner, Bandura u.a.; mit 42 Tabellen*, UTB für Wissenschaft Große Reihe Psychologie, Bd. 8035 (4., völlig neu bearb. Aufl.). München: Reinhardt.
- Pervin, L. A., Cervone, D. & John, O. P. (2005). *Persönlichkeitstheorien*. München: Ernst Reinhardt Verlag.
- Pintarich, H. E. (2002). "Lieb sein" als pädagogisches Konzept. In M. Sertl & B. Falkinger (Hrsg.), *LehrerInnenbildung in Bewegung? Zur Reform der Pädagogischen Akademien*, 181–189. Wien: Verein der Förderer der Schulhefte (= Schulheft 108).
- Polanyi, M. (1985). *Implizites Wissen*, Suhrkamp-Taschenbuch Wissenschaft, Bd. 543 (2. Auflage). Frankfurt am Main: Suhrkamp.
- Polk, J. A. (2006). Traits of Effective Teachers. *Arts Education Policy Review*, 107 (4), 23–29.
- Poropat, A. (2009). A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance. *Psychological Bulletin*, 135 (2), 322–338.

- Praetorius, A.-K., Pauli, C., Reusser, K., Rakoczy, K. & Klieme, E. (2014). One lesson is all you need? Stability of instructional quality across lessons. *Learning and Instruction, 31*, 2–12.
- Prave, R. S. & Baril, G. L. (1993). Instructor Ratings. Controlling for Bias from Initial Student Interest. *Journal of Education for Business, 68* (6), 362–366. <https://doi.org/10.1080/08832323.1993.10117644>
- Prenzel, M. & Allolio-Näcke, L. (Hrsg.). (2006). *Untersuchungen zur Bildungsqualität von Schule. Abschlussbericht des DFG-Schwerpunktprogramms*. Münster: Waxmann.
- Prenzel, M. & Baumert, J. (Hrsg.). (2009). *Vertiefende Analysen zu PISA 2006*. Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-91815-0>
- Raad, B. de. (1998). Five Big, Big Five Issues. Rationale, Content, Structure, Status, and Crosscultural Assessment. *European Psychologist, 3* (2), 113–124.
- Raad, B. de. (2000). *The big five personality factors. The psycholexical approach to personality*. Seattle: Hogrefe & Huber.
- Raad, B. de, Barelds, D. P. H., Levert, E., Ostendorf, F., Mlacic, B., Di Blas, L. et al. (2010). Only three factors of personality description are fully replicable across languages: a comparison of 14 trait taxonomies. *Journal of Personality and Social Psychology, 98* (1), 160–173. <https://doi.org/10.1037/a0017184>
- Radisch, F., Stecher, L., Klieme, E. & Kühnbach, O. (2007). Unterrichts- und Angebotsqualität aus Schülersicht. In H. G. Holtappels, E. Klieme, T. Rauschenbach & L. Stecher (Hrsg.), *Ganztagsschule in Deutschland. Ergebnisse der Ausgangserhebung der "Studie zur Entwicklung von Ganztagsschulen" (StEG)*, 227–260. Weinheim: Juventa.
- Raftery, A. E. (1993). Bayesian model selection in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models*, Sage focus editions, vol. 154, 163–180. Newbury Park, Calif.: Sage.
- Raftery, A. E. (1995). Bayesian Model Selection in Social Research. *Sociological Methodology, 25*, 111. <https://doi.org/10.2307/271063>
- Rahn, S., Keune, M., Fuhrmann, C. & Gruehn, S. (2016). Aus Schüleraussagen lernen?! Auf dem Weg zu einer professionellen Feedbackkultur an deutschen Schulen. *Die Deutsche Schule, 108* (2), 163–175. im Druck.
- Rakoczy, K., Buff, A. & Lipowsky, F. (2005). Erhebungsinstrumente. Teil 1. In E. Klieme, C. Pauli & K. Reusser (Hrsg.), *Dokumentation der Erhebungs- und Auswertungsinstrumente zur Schweizerisch- Deutschen Videostudie. Unterrichtsqualität, Lernverhalten und mathematisches Verständnis*, Materialien zur Bildungsforschung, Bd. 13. Frankfurt am Main: GFPF.

- Rakoczy, K., Buff, A. & Lipowsky, F. (2005). *Dokumentation der Erhebungs- und Auswertungsinstrumente zur schweizerisch-deutschen Videostudie. "Unterrichtsqualität, Lernverhalten und mathematisches Verständnis"* (Klieme, E., Pauli, C. & Reusser, K., Hrsg.). Frankfurt am Main.
- Rammsayer, T. & Weber, H. (2010). *Differentielle Psychologie - Persönlichkeitstheorien* (Bachelorstudium Psychologie). Göttingen: Hogrefe.
- Rammstedt, B., Koch, K., Borg, I. & Reitz, T. (2004). Entwicklung und Validierung einer Kurzsкала für die Messung der Big-Five-Persönlichkeitsdimensionen in Umfragen. *ZUMA Nachrichten*, 28 (55), 5–28.
- Ramsden, P. (1991). A performance indicator of teaching quality in higher education. The Course Experience Questionnaire. *Studies in Higher Education*, 16 (2), 129–150. <https://doi.org/10.1080/03075079112331382944>
- Randoll, D. (1995). *Schule im Urteil von Lehrern. Ergebnisse einer Befragung von Oberstufenlehrern aus vier alten und einem neuen Bundesland zu ihrer Wahrnehmung von Schule*, Studien zur Pädagogischen Psychologie, Bd. 32. Göttingen: Hogrefe Verl. für Psychologie.
- Rauthmann, J. F. (2017). *Persönlichkeitspsychologie. Paradigmen - Strömungen - Theorien* (Springer-Lehrbuch). Berlin, Heidelberg: Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-662-53004-7>
- Reinders, H. (Hrsg.). (2015). *Empirische Bildungsforschung. Gegenstandsbereiche* (2., überarb. Aufl.). Wiesbaden: Springer VS. <https://doi.org/10.1007/978-3-531-19994-8>
- Reinecke, J. (2005). *Strukturgleichungsmodelle in den Sozialwissenschaften*. München: Oldenbourg.
- Renaud, R. D. & Murray, H. G. (1996). Aging, personality, and teaching effectiveness in academic psychologists. *Research in Higher Education*, 37 (3), 223–240. <https://doi.org/10.1007/BF01730120>
- Rinderman, H. (2003). Lehrevaluation an Hochschulen: Schlussfolgerungen aus Forschung und Anwendung für Hochschulunterricht und seine Evaluation. *Zeitschrift für Evaluation*, 2, 233–256.
- Rinderman, H. & Oubaid, V. (1999). Auswahl von Studienanfängern. Vorschläge für ein zuverlässiges Verfahren. *Hochschulpolitik aktuell*, 11, 589-592.
- Rinderman, H. & Oubaid, V. (1999). Auswahl von Studienanfängern durch Universitäten-Kriterien. Verfahren und Prognostizierbarkeit des Studienerfolgs. *Zeitschrift für differentielle und diagnostische Psychologie*, 20 (3), 172–191.

- Roberts, B. W. & Wood, D. (2006). Personality Development in the Context of the Neo-Socioanalytic Model of Personality. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development*, 11–39. Mahwah, NJ: Erlbaum. Verfügbar unter <http://psycnet.apa.org/psycinfo/2006-07023-002>
- Roberts, B. W. & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age. A quantitative review of longitudinal studies. *Psychological Bulletin*, 126 (1), 3–25. <https://doi.org/10.1037/0033-2909.126.1.3>
- Robins, R. W., Fraley, R. C. & Krueger, R. F. (Eds.). (2007). *Handbook of research methods in personality psychology*. New York: Guilford Press.
- Robins, R. W., Fraley, R. C. & Krueger, R. F. (Hrsg.). (2007). *Handbook of Research Methods in Personality Psychology*. New York: The Guilford Press.
- Roeder, P. M. & Gruehn, S. (1997). Geschlecht und Kurswahlverhalten. *Zeitschrift für Pädagogik*, 43 (6), 877–894.
- Roloff Henoch, J., Klusmann, U., Lüdtke, O. & Trautwein, U. (2015). Who becomes a teacher? Challenging the “negative selection” hypothesis. *Learning and Instruction*, 36, 46–56. <https://doi.org/10.1016/j.learninstruc.2014.11.005>
- Rosenberg, M., Schooler, C., Schoenbach, C. & Rosenberg, F. (1995). Global self-esteem and specific self-esteem. Different concepts, different outcomes. *American Sociological Review*, 60 (1), 114–156.
- Rost, D. H. (Hrsg.). (2006). *Handwörterbuch Pädagogische Psychologie* (Schlüsselbegriffe, 3., überarb. und erw. Aufl.). Weinheim: Beltz PVU.
- Rowan, B., Correnti, R. & Miller, R. J. (November 2002). *What Large-Scale Survey Tells Us about Teacher Effects on Student Achievement. Insights from the "Prospects" Study of Elementary Schools*. CPRE Research Report Series. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63 (3), 581–592. <https://doi.org/10.1093/biomet/63.3.581>
- Rubin, R. B. & Feezel, J. D. (2009). Teacher communication competence. Essential skills and assessment procedures. *Central States Speech Journal*, 36 (1-2), 4–13. <https://doi.org/10.1080/10510978509363194>
- Rubin, A., & Bruce, B. C. (1985). *Learning with QUILL: Lessons for Students, Teachers and Software Designers* (National Inst. of Education (ED), Washington, DC., Hrsg.) (Reading Education Report No. 60.). Illinois Univ., Urbana. Center for the Study of Reading.; Bolt, Beranek and Newman, Inc., Cambridge, MA.
- Saldern, M. v., Littig, K. E. & Ingenkamp, K. (Hrsg.). (1986). *Landauer Skalen zum Sozialklima für 4. bis 13. Klassen (LASSO 4-13)*. Weinheim: Beltz.

- Saldern, M. v. (1987). *Sozialklima von Schulklassen. Überlegungen u. mehrebenenanalyt. Unters. zur subjektiven Wahrnehmung von Lernumwelten*, Europäische Hochschulschriften Reihe 11, Bd. 302. Zugl.: Mainz, Erziehungswiss. Hochsch., Diss., 1986. Frankfurt am Main: Lang.
- Salgado, J. F. (1997). The five factor model of personality and job performance in the European Community. *Journal of Applied Psychology*, 82 (1), 30–43. <https://doi.org/10.1037/0021-9010.82.1.30>
- Sarges, W. & Wottawa, H. (Hrsg.). (2001). *Handbuch wirtschaftspsychologischer Testverfahren*. Lengerich: Pabst Science Publishers.
- Sauter, F. C. (Hrsg.). (2003). *Schulwirklichkeit und Wissenschaft. Ausgewählte Kongressbeiträge von Didaktikern, Pädagogen, Psychologen*, Schriftenreihe Erziehung - Unterricht - Bildung, Bd. 101. Hamburg: Kovač.
- Sauter, F. C. (2006). *Das autonome Erziehungsobjekt. Funktion oder Ziel der Pädagogik?* Dissertation, Universität Hagen. Hagen.
- Schaal, B. (Hrsg.). (2010). *Qualitätssicherung im Bildungswesen. Auftrag und Anspruch der bayerischen Qualitätsagentur; eine Publikation des Staatsinstituts für Schulqualität und Bildungsforschung (ISB)*. Münster: Waxmann.
- Schaarschmidt, U. (2001). Psychische Gesundheit im Lehrerberuf. In N. Seibert (Hrsg.), *Probleme der Lehrerbildung. Analysen, Positionen, Lösungsversuche*, 69–92. Bad Heilbrunn/Obb.: Klinkhardt.
- Schaarschmidt, U. & Fischer, A. W. (2001). *Bewältigungsmuster im Beruf. Persönlichkeitsunterschiede in der Auseinandersetzung mit der Arbeitsbelastung; mit 86 Abbildungen und 6 Tabellen*. Göttingen: Vandenhoeck & Ruprecht.
- Schaarschmidt, U. & Fischer, A. W. (2008). *AVEM. Arbeitsbezogenes Verhaltens- und Erlebensmuster* (3. überarbeitete und erweiterte Auflage). Frankfurt am Main: Pearson. Verfügbar unter pearsonassessment.de
- Schaarschmidt, U. & Fischer, A. W. (2013). *Lehrergesundheit fördern - Schulen stärken. Ein Unterstützungsprogramm für Kollegium und Leitung* (Pädagogik 2014). Weinheim: Beltz.
- Schäfer, A. (2005). *Einführung in die Erziehungsphilosophie*. Weinheim: Beltz.
- Schaffhuser, K. (2014). *Personality and Its Inside and Outside: A Dyadic, Developmental Perspective*. University of Zurich. Zurich.
- Schaffhuser, K., Allemand, M. & Martin, M. (2014). Personality Traits and Relationship Satisfaction in Intimate Couples. Three Perspectives on Personality. *European Journal of Personality*, 28 (2), 120–133. <https://doi.org/10.1002/per.1948>

- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the Fit of Structural Equation Models. Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of psychological research online*, 8 (2), 23–74.
- Schiefele, U. (1991). Interest, learning, and motivation. *Educational Psychologist*, 26 (3 & 4), 299–323.
- Schiefele, U. (1991). Interest, Learning, and Motivation. *Educational Psychologist*, 26 (3-4), 299–323. <https://doi.org/10.1080/00461520.1991.9653136>
- Schiefele, U. (1996). Topic Interest, Text Representation, and Quality of Experience. *Contemporary Educational Psychology*, 21 (1), 3–18. <https://doi.org/10.1006/ceps.1996.0002>
- Schiefele, U. (Hrsg.). (2000). *Interesse und Lernmotivation. Untersuchungen zu Entwicklung, Förderung und Wirkung; [Andreas Krapp zu seinem 60. Geburtstag am 3. Juli 2000 gewidmet]*. Münster: Waxmann.
- Schiefele, U. (2009). Situational and individual interest. In K. R. Wentzel & A. Wigfield (Eds.), *Handbook of motivation at school*, Educational psychology handbook series, 197–222. New York: Routledge.
- Schiefele, U., Krapp, A. & Schreyer, I. (1993). Metaanalyse des Zusammenhangs von Interesse und schulischer Leistung. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 25 (2), 120–148.
- Schmidt, C. P., Lewis, B. E. & Kurpius-Brock, M. J. (1991). Relationships Between Teacher Personality and Ratings of Applied Music Teaching Behavior. *Contributions to Music Education*, 20–35.
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L. & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, 80 (5), 607–620. <https://doi.org/10.1037/0021-9010.80.5.607>
- Schrader, F. W. & Helmke, A. (2002). Motivation, Lernen und Leistung. In A. Helmke & R. S. Jäger (Hrsg.), *Das Projekt MARKUS. Mathematik-Gesamterhebung Rheinland-Pfalz: Kompetenzen, Unterrichtsmerkmale, Schulkontext*, 257–324. Landau: Verl. Empirische Pädagogik e.V.
- Schrader, F.-W. & Helmke, A. (2008). Determinanten der Schulleistung. In M. K. W. Schweer (Hrsg.), *Lehrer-Schüler-Interaktion*, 285–302. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91104-5_11
- Schröder, S. (1999, 1. Februar). *Konstrukte der Unterrichtsqualität im Expertenurteil*. Unveröffentlichte Diplomarbeit, Freie Universität Berlin. Berlin.
- Schulz von Thun, F. (2005). *Miteinander reden*, rororo Sachbuch, Bd. 61151, Sonderausg, 3 Bände. Hamburg: Rowohlt.

- Schumann, S. (Hrsg.). (2005). *Persönlichkeit. Eine vergessene Größe der empirischen Sozialforschung* (1. Aufl.). Wiesbaden: VS Verl. für Sozialwiss.
- Schumann-Hengsteler, R. (Hrsg.). (1996). *Entwicklung im Jugendalter*. Göttingen: Hogrefe Verl. für Psychologie.
- Schupp, J. & Gerlitz, J.-Y. (2005). *Zur Erhebung der Big-Five-basierten Persönlichkeitsmerkmale im SOEP. Dokumentation der Instrumententwicklung BFI-S auf Basis des SOEP-Pretests 2005*. Berlin: Deutsches Institut für Wirtschaftsforschung.
- Schupp, J. & Gerlitz, J.-Y. (2008). *Informationen zum Instrument "Big Five Inventory-SOEP (BFI-S; Schupp & Gerlitz)"*. Kurzbericht.
- Schurig, M., Glesemann, B. & Schröder, J. (2016). Dimensionen von Unterrichtsqualität. Die Generalisierbarkeit von Schülerurteilen über Fächer und Zeit. In R. Strietholt, W. Bos, H. G. Holtappels & N. McElvany (Hrsg.), *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven, Bd. 19*, 30–56. Weinheim: Beltz Juventa.
- Schwanenberg, J., Winkelsett, D. & Schurig, M. (2015). Methoden der wissenschaftlichen Begleitforschung im Projekt Ganz. In H. Wendt & W. Bos (Hrsg.), *Auf dem Weg zum Ganztagsgymnasium. Erste Ergebnisse der wissenschaftlichen Begleitforschung zum Projekt Ganz In*, 493–516. Münster: Waxmann.
- Schwarz, B. & Prange, K. (Hrsg.). (1997). *Schlechte Lehrer/innen. Zu einem vernachlässigten Aspekt des Lehrerberufs*. Weinheim: Beltz.
- Schwarz, B. & Prange, K. (Hrsg.). (1997). *Schlechte Lehrer/innen. Zu einem vernachlässigten Aspekt des Lehrberufs* (Beltz grüne Reihe). Weinheim: Beltz.
- Schweer, M. K. W. (Hrsg.). (2008). *Lehrer-Schüler-Interaktion*. Wiesbaden: VS Verlag für Sozialwissenschaften. <https://doi.org/10.1007/978-3-531-91104-5>
- Schweer, M. K. W. (2008). *Lehrer-Schüler-Interaktion. Inhaltsfelder, Forschungsperspektiven und methodische Zugänge*, Schule und Gesellschaft, Bd. 24 (2., vollständig überarbeitete Auflage). Wiesbaden: VS Verlag für Sozialwissenschaften / GWV Fachverlage GmbH Wiesbaden. <https://doi.org/10.1007/978-3-531-91104-5>
- Scollon, C. N. & Diener, E. (2006). Love, work, and changes in extraversion and neuroticism over time. *Journal of Personality and Social Psychology*, 91 (6), 1152–1165. <https://doi.org/10.1037/0022-3514.91.6.1152>
- Scriven, M. (1981). Summative Teacher Evaluation. In J. Millman (Ed.), *Handbook of teacher evaluation*, 244–271. Beverly Hills: Sage.
- Seibert, N. (Hrsg.). (2001). *Probleme der Lehrerbildung. Analysen, Positionen, Lösungsversuche*. Bad Heilbrunn/Obb.: Klinkhardt.

- Seidel, T. & Shavelson, R. (2007). Teaching Effectiveness Research in the Past Decade. The Role of Theory and Research Design in Disentangling Meta-Analysis Results. *Review of Educational Research*, 77 (4), 454–499.
- Seifried, J. (2006). Sichtweisen auf die methodische Gestaltung von Unterricht. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 102 (4), 578–597.
- Seitz, S. (2008). Der Lehrer und seine Persönlichkeit - Determinanten eines wiederentdeckten Begriffs. *Lehren und Lernen / Schriftenreihe* (12), 31–37.
- Sertl, M. & Falkinger, B. (Hrsg.). (2002). *LehrerInnenbildung in Bewegung? Zur Reform der Pädagogischen Akademien*. Wien: Verein der Förderer der Schulhefte (= Schulheft 108).
- Shayer, M. & Adhami, M. (2007). Fostering Cognitive Development Through the Context of Mathematics. Results of the CAME Project. *Educational Studies in Mathematics*, 64 (3), 265–291. <https://doi.org/10.1007/s10649-006-9037-1>
- Sherman, B. R. & Blackburn, R. T. (1974). Personal Characteristics and Teaching Effectiveness of College Faculty. *Journal of Educational Psychology*, 67, 124–131.
- Sieland, B. & Reißland, B. (Hrsg.). (2000). *Qualitätssicherung in der Lehrerbildung. Lehrerarbeit: Bedingungsfaktoren und Qualitätskriterien* (Studien zur Erwachsenenbildung, Bd. 9). Hamburg: Kovač.
- Silbereisen, R. K. (1996). Jugendliche als Gestalter ihrer Entwicklung. Konzepte und Forschungsbeispiele. In R. Schumann-Hengsteler (Hrsg.), *Entwicklung im Jugendalter*, 1–18. Göttingen: Hogrefe Verl. für Psychologie.
- Silbereisen, R. K. & Todt, E. (Hrsg.). (1994). *Adolescence in context. The interplay of family, school, peers, and work in adjustment*. New York: Springer-Verlag.
- Simms, L.J., Zelazny, K., Yam, W.H., Gros, D.F. (2010). Self-informant agreement for personality evaluative person descriptors: Comparing methods for creating informant measures. *European Journal of Personality*, 24, 207-221.
- Simon, A. (2006). *Dimensionale Persönlichkeitstheorien. The Big Five*. München: Grin Verlag.
- Slavin, R. E. (1994). Quality, appropriateness, incentive, and time. A model of instructional effectiveness. *International Journal of Educational Research*, 21 (2), 141–157. [https://doi.org/10.1016/0883-0355\(94\)90029-9](https://doi.org/10.1016/0883-0355(94)90029-9)
- Small, A., Hollenbeck, A. R. & Haley, R. L. (1982). The effect of emotional state on student ratings of instructors. *Teaching of Psychology*, 9, 205–211.
- Smart, J. C. (Hrsg.). (1992). *Higher Education: Handbook of Theory and Research* (8. Aufl.). Netherlands: Springer.
- Solomon, D. (1966). Teacher behavior dimensions, course characteristics, and student evaluations of teachers. *American Educational Research Journal*, 3 (1), 35–47.

- Sorey, K. E. (1968). A study of the distinguishing personality characteristics of college faculty who are superior in regard to the teaching function. *Dissertation Abstracts International*, 28 (12-A).
- Specht, J., Egloff, B. & Schmukle, S. C. (2011). *Stability and Change of Personality Across the Life Course. The Impact of Age and Major Life Events on Mean-Level and Rank-Order Stability of the Big Five* (Meran, G., Wagner, G. G., Frick, J. R., Schupp, J., D'Ambrosio, C. & Breuer Christoph, et al., Hrsg.). Berlin: Deutsches Institut für Wirtschaftsforschung.
- Spencer, P. A. (1992). *Improving Teacher Evaluation*. Riverside, CA: Riverside Community College.
- Spencer, P. A. & Flyr, M. L. (1992). *The formal evaluation as an impetus to classroom change: Myth or reality?* Riverside, CA: CA Research/Technical Report.
- Spiel, C. & Gössler, P. M. (2000). Zum Einfluss von Biasvariablen auf die Bewertung universitärer lehre durch Studierende. *Zeitschrift für Pädagogische Psychologie*, 14 (1), 38–47.
- Spiewak, M. Ich bin superwichtig! *Zeit online*. Verfügbar unter <http://www.zeit.de/2013/02/Paedagogik-John-Hattie-Visible-Learning>
- Spinath, B. (2015). Lernmotivation. In H. Reinders (Hrsg.), *Empirische Bildungsforschung. Gegenstandsbereiche* (2., überarb. Aufl.), 55–67. Wiesbaden: Springer VS.
- Srivastava, S. (2012). Other People as a Source of Self-Knowledge. In S. Vazire & T. D. Wilson (Eds.), *Handbook of self-knowledge*, 91–104. New York, NY: Guilford Press.
- Srivastava, S., John, O. P., Gosling, S. D. & Potter, J. (2003). Development of personality in early and middle adulthood. Set like plaster or persistent change? *Journal of Personality and Social Psychology*, 84 (5), 1041–1053. <https://doi.org/10.1037/0022-3514.84.5.1041>
- Stapleton, L. M. (2016). Stapleton, L. M., Harring, J. R., & Lee, D. (2015). Sampling weight considerations for multilevel modeling of panel data. In J. Harring, L. M. Stapleton & S. N. Beretvas (Eds.), *Advances in multilevel modeling for educational research. Addressing practical issues found in real-world applications*, CILVR series on latent variable methodology, 63–96. Charlotte, NC: Information Age Publishing Inc.
- Stapleton, L. M., Yang, J. S. & Hancock, G. R. (2016). Construct Meaning in Multilevel Settings. *Journal of Educational and Behavioral Statistics*, 41 (5), 481–520. <https://doi.org/10.3102/1076998616646200>
- Steiger, J. H. & Lind, J. C. (1980). *Statistically Based Tests for the Number of Common Factors*, Annual Meeting of the Psychometric Society, 424–453. Iowa City, IA
- Sternfeld, N. (2009). *Das pädagogische Unverhältnis. Lehren und Lernen bei Rancière, Gramsci und Foucault*. Wien: Turia + Kant.
- Stoeckli, G. (2000). Burnout. Vom psychohygienischen zum pädagogischen Konstrukt. *Beiträge zur Lehrerbildung*, 18 (2), 199–206.

- Stolz, G. (1997). Der schlechte Lehrer aus der Sicht von Schülern. In B. Schwarz & K. Prange (Hrsg.), *Schlechte Lehrer/innen. Zu einem vernachlässigten Aspekt des Lehrerberufs*. Weinheim: Beltz.
- Strietholt, R., Bos, W., Holtappels, H. G. & McElvany, N. (Hrsg.). (2016). *Jahrbuch der Schulentwicklung. Daten, Beispiele und Perspektiven* (Bd. 19). Weinheim: Beltz Juventa.
- Sweet, A. P., Guthrie, J. T. & Ng, M. M. (1998). Teacher perceptions and student reading motivation. *Journal of Educational Psychology, 90* (2), 210–223. <https://doi.org/10.1037/0022-0663.90.2.210>
- Tatro, C. N. (1995). Gender effects on student evaluations of faculty. *Journal of Research & Development in Education, 28* (3), 169–173. Verfügbar unter <http://psycnet.apa.org/psycinfo/1995-45371-001>
- Teml, H. & Unterweger, E. (2002). Persönlichkeitsförderung in der LehrerInnenbildung. *Journal für Lehrerinnen- und Lehrerbildung, 2* (2), 7–21.
- Terhart, E. (1997). Guter Lehrer - schlechter Lehrer: Zur Frage der Beurteilung der Arbeitsleistung von Lehrkräften. In B. Schwarz & K. Prange (Hrsg.), *Schlechte Lehrer/innen. Zu einem vernachlässigten Aspekt des Lehrerberufs*, 34–85. Weinheim: Beltz.
- Terhart, E. (2000). Qualität und Qualitätssicherung im Schulsystem. Hintergründe - Konzepte - Probleme. *Zeitschrift für Pädagogik, 26* (6), 809–829.
- Terhart, E. (2006). Was wissen wir über gute Lehrer? *Pädagogik* (5), 42–47.
- Terhart, E. (2007). Was wissen wir über gute Lehrer? Ergebnisse aus der empirischen Lehrerforschung. *Friedrich Jahresheft, 25*, 20–24.
- Terhart, E. (2012). Was wissen wir über Lehrerinnen und Lehrer? Eine Einführung in die Serie. *Pädagogik, 64* (1), 43–47.
- Terhart, E., Bennewitz, H. & Rothland, M. (Hrsg.). (2011). *Handbuch der Forschung zum Lehrerberuf*. Münster: Waxmann.
- Todt, E. (1995). Entwicklung des Interesses. In H. Hetzer & R. Arbinger (Hrsg.), *Angewandte Entwicklungspsychologie des Kindes- und Jugendalters*, UTB für Wissenschaft, Bd. 935 (3. Aufl., unveränd. Nachdr. der 2., überarb. Aufl.), 213–264. Heidelberg: Quelle & Meyer.
- Todt, E., Drewes, R. & Heils, S. (1994). The development of interests during adolescence: Social context, individual differences, and individual significance. In R. K. Silbereisen & E. Todt (Hrsg.), *Adolescence in context. The interplay of family, school, peers, and work in adjustment*, 82–95. New York: Springer-Verlag.
- Todt, E. & Schreiber, S. (1998). Development of interests. In L. Hoffmann (Ed.), *Interest and learning. Proceedings of the Seeon Conference on Interest and Gender*, IPN, vol. 164, 25–40. Kiel: IPN.

- Tokar, D. M., Fischer, A. R. & Mezydlo Subich, L. (1998). Personality and Vocational Behavior. A Selective Review of the Literature, 1993–1997. *Journal of Vocational Behavior*, 53 (2), 115–153. <https://doi.org/10.1006/jvbe.1998.1660>
- Tomasco, A. T. (1980). Student Perceptions of Instructional and Personality Characteristics of Faculty: A Canonical Analysis. *Teaching of Psychology*, 7 (2), 79–82. Verfügbar unter http://www.tandfonline.com/doi/abs/10.1207/s15328023top0702_3?journalCode=htop20
- Trautwein, U. & Lüdtke, O. (2004). Aspekte von Wissenschaftspropädeutik und Studierfähigkeit. In O. Köller, R. Watermann, U. Trautwein & O. Lüdtke (Hrsg.), *Wege zur Hochschulreife in Baden-Württemberg. TOSCA - Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien*, 327–366. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Treumann K. P., Neubauer G., Möller R. & Abel J. (Hrsg.). (1996). *Methoden und Anwendungen empirischer pädagogischer Forschung*. Münster und New York: Waxmann.
- Tsai, Y.-M., Kunter, M., Lüdtke, O., Trautwein, U., Ryan, R. & Ryan, R. M. (2008). What Makes Lessons Interesting? // What makes lessons interesting? The Role of Situational and Individual Factors in Three School Subjects // The role of situational and individual factors in three school subjects. *Journal of Educational Psychology*, 100 (2), 460–472. <https://doi.org/10.1037/0022-0663.100.2.460>
- Unbekannt (Dudenverlag, Hrsg.). *Wörterbucheintrag zu "Persönlichkeit"*, Bibliographisches Institut GmbH. Zugriff am 23.06.2017. Verfügbar unter <http://www.duden.de/rechtschreibung/Persoenlichkeit>
- Unbekannt (Dudenverlag, Hrsg.). (2016, 24. Juni). *Wörterbucheintrag zu "Identität"*, Bibliographisches Institut GmbH. Verfügbar unter <http://www.duden.de/rechtschreibung/Identitaet>
- Universität Passau (Hrsg.). *Über PArcours*, Philosophische Fakultät - Lehrstuhl für Schulpädagogik. Verfügbar unter <http://www.phil.uni-passau.de/schulpaedagogik/forschungprofilelemente/parcours/ueber-parcours/>
- Unkelbach, C. & Memmert, D. (2014). Serial-Position Effects in Evaluative Judgments. *Current Directions in Psychological Science*, 23 (3), 195–200. <https://doi.org/10.1177/0963721414533701>
- Unterweger, E. (2014). *Personenbezogene überfachliche Kompetenzen von Pädagoginnen und Pädagogen. Selbst-, Sozial- und Systemkompetenzen*.
- Urban, D. & Mayerl, J. (2011). *Mediator-Effekte in der Regressionsanalyse (direkte, indirekte und totale Effekte)* (Version 1.3).
- Urban, W. (1984). *Persönlichkeitsstruktur und Unterrichtskompetenz*, Lehrerbildung - aktuell, Bd. 4. Wien: Österr. Bundesverl.

- Vazire, S. (2010). Who knows what about a person? The self-other knowledge asymmetry (SOKA) model. *Journal of Personality and Social Psychology*, 98 (2), 281–300. <https://doi.org/10.1037/a0017908>
- Vazire, S. & Carlson, E. N. (2010). Self-Knowledge of Personality. Do People Know Themselves? *Social and Personality Psychology Compass*, 4 (8), 605–620. <https://doi.org/10.1111/j.1751-9004.2010.00280.x>
- Vazire, S. & Wilson, T. D. (Eds.). (2012). *Handbook of self-knowledge*. New York, NY: Guilford Press.
- Institut für Demoskopie Allensbach. (2012, 24. Oktober). *Lehre(r) in Zeiten der Bildungspolitik. Eine Studie zum Prestige des Lehrerberufs und zur Situation an den Schulen in Deutschland* (Vodafone Stiftung Deutschland, Hrsg.).
- Vroeijenstijn, T. (1992). External quality assessment, servant of two masters? The Netherlands university perspective. In A. Craft (Ed.), *Quality assurance in higher education. Proceedings of an international conference Hong Kong, 1991*, 109–131. London: Falmer Press.
- Wachtel, H. K. (1998). Student Evaluation of College Teaching Effectiveness. A brief review. *Assessment & Evaluation in Higher Education*, 23 (2), 191–211.
- Wagenaar, T. C. (1995). Student Evaluation of Teaching. Some Cautions and Suggestions. *Teaching Sociology*, 23 (1), 64–68.
- Wagner, W. (2008). *Methodenprobleme bei der Analyse der Unterrichtswahrnehmung und-wirksamkeit—am Beispiel der Studie DESI (Deutsch Englisch Schülerleistungen International) der Kultusministerkonferenz*. Dissertation, Universität Koblenz-Landau. Landau.
- Wagner, W., Göllner, R., Helmke, A., Trautwein, U. & Lüdtke, O. (2013). Construct validity of student perceptions of instructional quality is high, but not perfect. Dimensionality and generalizability of domain-independent assessments. *Learning and Instruction*, 28, 1–11. <https://doi.org/10.1016/j.learninstruc.2013.03.003>
- Waldis, M., Buff, A., Pauli, C. & Reusser, K. (2002). *Skalendokumentation zur Schülerinnen- und Schülerbefragung im schweizerischen Videoprojekt*, Universität Zürich.
- Waldis, M. (2012). *Interesse an Mathematik. Zum Einfluss des Unterrichts auf das Interesse von Schülerinnen und Schülern der Sekundarstufe I*. Münster: Waxmann.
- Walter, O. B. & Rose, M. (2013). Effect of item order on item calibration and item bank construction for computer adaptive tests. In U. Kröhne & A. Frey (Hrsg.), *Psychological Test and Assessment Modeling*, Volume 55, Issue 1. Pabst Science Publishers.
- Wang, M., Haertel, G. & Walberg, H. (1993). Synthesis of Research. What Helps Students Learn? *Educational Leadership*, 51 (4), 74–79.

- Watson, D., Hubbard, B. & Wiese, D. (2000). Self-other agreement in personality and affectivity. The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology*, 78 (3), 546–558. <https://doi.org/10.1037//0022-3514.78.3.546>
- Weber, H. & Rammsayer, T. (2012). *Differentielle Psychologie. Persönlichkeitsforschung*. Göttingen: Hogrefe Verlag.
- Weiber, R. & Mülhhaus, D. (2010). *Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS; [Extras im Web]* (Springer-Lehrbuch). Heidelberg: Springer.
- Weinert, F. E. (1997). Notwendige Methodenvielfalt. Unterschiedliche Lernfähigkeiten erfordern variable Unterrichtsmethoden. *Friedrich Jahresheft*, 50–52.
- Weinert, F. E. & Helmke, A. (1996). Der gute Lehrer. Person, Funktion oder Fiktion? In A. Leschinsky (Hrsg.), *Die Institutionalisierung von Lehren und Lernen. Beiträge zu einer Theorie der Schule; Festgabe für Peter Martin Roeder*, 223–233. Weinheim: Beltz.
- Weinert, F. E. & Helmke, A. (Hrsg.). (1997). *Entwicklung im Grundschulalter*. Weinheim: Beltz Psychologie-Verl.-Union.
- Weiß, P., Winkler, A., Zoli-Sudbrock, R. & Feller Mayer, M. (Hrsg.). (2010). *Langenscheidt Universal-Wörterbuch Latein. Lateinisch-deutsch, deutsch-lateinisch* (Neubearb.). Berlin: Langenscheidt.
- Weller, J. & Watson, D. (2009). Friend or foe? Differential use of the self-based heuristic as a function of relationship satisfaction. *Journal of personality*, 77 (3), 731–760. <https://doi.org/10.1111/j.1467-6494.2009.00563.x>
- Wendt, H. & Bos, W. (Hrsg.). (2015). *Auf dem Weg zum Ganztagsgymnasium. Erste Ergebnisse der wissenschaftlichen Begleitforschung zum Projekt Ganz In*. Münster: Waxmann.
- Wentzel, K. R. & Wigfield, A. (Eds.). (2009). *Handbook of motivation at school* (Educational psychology handbook series). New York: Routledge.
- Wenzel, M. (2011). Instrumente der Arbeits- und Organisationspsychologie. START-P: Testbatterie für Berufseinsteiger – Persönlichkeit. *Zeitschrift für Arbeits- und Organisationspsychologie*, 55 (3), 158–163.
- Wild, E. & Möller, J. (Hrsg.). (2009). *Pädagogische Psychologie* (Springer-Lehrbuch). Berlin, Heidelberg: Springer-Verlag Berlin Heidelberg. <https://doi.org/10.1007/978-3-540-88573-3>
- Willems, A. (2011). *Bedingungen des situationalen Interesses im Mathematikunterricht. Eine mehrebenenanalytische Perspektive*. Münster: Waxmann.
- Willems, A. & Glesemann, B. (2015). Individuelle Förderung und der Umgang mit Heterogenität im Fachunterricht an Ganztagsgymnasien. In H. Wendt & W. Bos (Hrsg.), *Auf dem Weg zum Ganztagsgymnasium. Erste Ergebnisse der wissenschaftlichen Begleitforschung zum Projekt Ganz In*, 414–443. Münster: Waxmann.

- Wirtz, M. A. (Hrsg.). (2014). *Dorsch. Lexikon der Psychologie* (17. Aufl.). Bern: Hans Huber.
- Wittwer, J. (2009). What Influences the Agreement Among Student Ratings of Science Instruction? In M. Prenzel & J. Baumert (Hrsg.), *Vertiefende Analysen zu PISA 2006*, 205–220. Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-531-91815-0_11
- Wolfer, T. & Johnson, M. (2003). Re-evaluating Student Evaluation of Teaching. The Teaching Evaluation Form. *Journal of Social Work Education*, 39 (1), 111–121. Verfügbar unter <http://www.jstor.org/stable/23043634>
- Wooldridge, J. M. (2013). *Introductory econometrics. A modern approach* (5. ed., internat. ed.). Mason Ohio: South-Western Cengage Learning.
- Young, S. & Shaw, D. G. (1999). Profiles of Effective College and University Teachers. *The Journal of Higher Education*, 70 (6), 670–686.
- Zinnbauer, M. & Eberl, M. (2004). *Überprüfung der Spezifikation und Güte von Strukturgleichungsmodellen. Verfahren und Anwendung*. Schriften zur Empirischen Forschung und Unternehmensplanung der Ludwig-Maximilians-Universität München, Heft 21/2004. <https://doi.org/10.15358/0340-1650-2005-10-566>
- Zlatkin-Troitschanskaia, O. (Hrsg.). (2009). *Lehrprofessionalität. Bedingungen, Genese, Wirkungen und ihre Messung* (Beltz-Bibliothek). Weinheim: Beltz.

Anhang

A: Fragebogen.....	305
B: Skalenhandbuch	309
C: Faktorladungen der konfirmatorischen Faktorenanalysen der selbsteingeschätzten Lehrer-BigFive – entlang BFI-S/gekürzt (N = 103).....	326
D: Korrelationen zwischen Persönlichkeitswahrnehmungen nach Geschlecht	327
E: Modelle (1a/2a/2b) – signifikante Regressionskoeffizienten der Lehrerpersönlichkeit auf die Unterrichtsqualität.	330
F: Werte der Gesamtkonstrukte Persönlichkeit und Unterrichtsqualität auf Faktoren 2. Ordnung und ihre Regressionseffekte	332

A: Fragebogen

Schülerbefragung

Lehrerpersönlichkeit: Wie schätzen Sie Ihre/n Deutsch-/Mathematiklehrer/in ein?

	1	2	3	4	5	6	7
	trifft überhaupt nicht zu						
	trifft voll zu						
Unsere Lehrerin/ unser Lehrer...	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...ist zurückhaltend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...ist kommunikativ, gesprächig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...ist eher faul.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...geht aus sich heraus, ist gesellig.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...hat eine lebhaft Phantasie, Vorstellung.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...schätzt künstlerische, ästhetische Erfahrungen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...ist originell, bringt neue Ideen rein.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...arbeitet gründlich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...erledigt Aufgaben wirksam und effizient.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...macht sich oft Sorgen.*	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...ist entspannt, kann mit Stress gut umgehen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...wird leicht nervös.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...ist manchmal etwas grob zu anderen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...geht rücksichtsvoll und freundlich mit anderen um.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...kann verzwehlen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

* wurde nach eingehender Validitätsprüfung aus Skala entfernt

Unterrichtsqualität (Motivierung): Wie werden Sie in Ihrem Deutsch-/Mathematikunterricht motiviert?

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft zu
Unsere Lehrerin/unsere Lehrerin gestaltet den Unterricht oft sehr spannend.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Unsere Lehrerin/unsere Lehrerin kann auch trockenen Stoff wirklich interessant machen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Unsere Lehrerin/unsere Lehrerin kann Schüler/innen manchmal richtig begeistern.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Unterrichtsqualität (Alltagsbezug): Wie werden Sie in Ihrem Deutsch-/Mathematikunterricht motiviert?

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft zu
Unsere Lehrerin/unsere Lehrerin zeigt uns an Beispielen aus dem täglichen Leben, wozu man das Gelernte gebrauchen kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Wenn wir in diesem Unterricht etwas Neues erarbeiten, gehen wir meistens von unseren eigenen Erfahrungen und Alltagsbeispielen aus.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Um uns etwas zu erklären, nimmt unsere Lehrerin/unsere Lehrerin oft ein Beispiel aus dem täglichen Leben.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Unterrichtsqualität (Angemessene Methodenvariation): Wie denken Sie über die methodische Gestaltung Ihres Deutsch-/Mathematikunterrichts?

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft zu
In diesem Mathematikunterricht findet häufig eine Variation der Methoden statt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Dieser Mathematikunterricht ist methodisch abwechslungsreich.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
In diesem Mathematikunterricht werden immer dieselben Methoden eingesetzt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Unterrichtsqualität (Disziplin): Wie denken Sie über die Art und Weise, wie Ihre Klasse in Ihrem Deutsch-/Mathematikunterricht geführt wird?

	trifft nicht zu	trifft eher nicht zu	trifft eher zu	trifft zu
In diesem Mathematikunterricht dauert es meistens sehr lange, bis alle zur Arbeit bereit sind.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
In diesem Mathematikunterricht wird oft sehr gestört.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
In diesem Mathematikunterricht fehlt meistens bei irgend jemandem etwas, wenn wir anfangen wollen zu arbeiten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Unterrichtsqualität (Störungsprävention): Wie denken Sie über die Art und Weise, wie Ihre Klasse in Ihrem Deutsch-/Mathematikunterricht geführt wird?			
	trifft nicht zu	trifft eher nicht zu	trifft eher zu
In diesem Mathematikunterricht hat unsere Lehrerin/ unser Lehrer klargemacht, was passiert, wenn man (Verhaltens-) Regeln verletzt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
In diesem Mathematikunterricht wird klar, was man machen darf und was nicht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Unsere Lehrerin/ unser Lehrer merkt sofort, wenn Schüler/innen beginnen, sich mit etwas anderem zu beschäftigen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Unsere Lehrerin/ unser Lehrer achtet sehr darauf, dass wir aufpassen.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Unterrichtsqualität (Erklärkompetenz): Wie denken Sie über die Klarheit und Struktur Ihres Deutsch-/ Mathematikunterrichts?			
	trifft nicht zu	trifft eher nicht zu	trifft eher zu
Unsere Lehrerin/ unser Lehrer hebt immer wieder hervor, was wichtig ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Unsere Lehrerin/ unser Lehrer kann gut erklären.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Unsere Lehrerin/ unser Lehrer fasst häufig nochmals den Stoff zusammen, damit wir ihn uns gut merken können.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Unsere Lehrerin/ unser Lehrer erklärt so, dass keiner weiß, was los ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Unterrichtsqualität (Angemessenes Interaktionstempo): Wie denken Sie über das Unterrichtstempo und die Anforderungen in Ihrem Deutsch-/ Mathematikunterricht?			
	trifft nicht zu	trifft eher nicht zu	trifft eher zu
Unsere Lehrerin/ unser Lehrer lässt bei Fragen kaum Zeit zum Nachdenken.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Unsere Lehrerin/ unser Lehrer verlangt oft blitzschnelle Antworten.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Unsere Lehrerin/ unser Lehrer geht gleich zum Nächsten, wenn nicht sofort geantwortet wird.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Lehrerbefragung

Lehrerpersönlichkeit: Wie schätzen Sie sich persönlich ein?

	trifft überhaupt nicht zu							trifft voll zu
Ich bin ein Mensch, der...	1	2	3	4	5	6	7	7
...zurückhaltend ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...kommunikativ, Gesprächig ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...eher faul ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...aus sich herausgeht, gesellig ist.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...eine lebhaft Phantasie, Vorstellung hat.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...künstlerische, ästhetische Erfahrungen schätzt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...originell ist, neue Ideen reinbringt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...gründlich arbeitet.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...Aufgaben wirksam und effizient erledigt.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...sich oft Sorgen macht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...entspannt ist, mit Stress gut umgehen kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...leicht nervös wird.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...manchmal etwas grob zu anderen ist. *	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...rücksichtsvoll und freundlich mit anderen umgeht.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
...verzeihen kann.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

* wurde nach eingehender Validitätsprüfung aus Skala entfernt

B: Skalenhandbuch

Schülererhebung

Persönlichkeitsdimensionen

Gewissenhaftigkeit

Kurzbezeichnung: D14, M14

Literatur Schupp, J., & Gerlitz, J.-Y. (2008). BFI-S: Big Five Inventory – SOEP. In A. Glockner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 12.00. Bonn: GESIS.

Anzahl Items: 3

Skalierung: 1=„trifft überhaupt nicht zu“; 7=„trifft voll zu“

Wie schätzen Sie Ihre/n Lehrer/in Ihres Deutschunterrichts/Mathematikunterrichts ein?

Variable		Skalenkennwerte			
Itemkürzel	Text	m	SD	rit	λ
<i>rec_D14c</i>	<i>Unsere Lehrerin/unsere/r Lehrer ist eher faul.</i>	5,907	1,512	0,506	0,736
<i>rec_M14c</i>		5,913	1,551	0,470	0,707
D14h	Unsere Lehrerin/unsere/r Lehrer arbeitet gründlich.	5,586	1,483	0,747	0,906
M14h		5,661	1,601	0,739	0,790
D14i	Unsere Lehrerin/unsere/r Lehrer erledigt Aufgaben wirksam und effizient.	5,432	1,575	0,730	0,901
M14i		5,556	1,614	0,707	0,895

Cronbach-Alpha(D)=0,807

Cronbach-Alpha (M)= 0,791

Skalenmittelwert (D)= 5,631

Skalenmittelwert (M)= 5,703

ICC (D)=0,259

ICC(M)=0,320

Standardabweichung (D)= 1,302

Standardabweichung (M)= 1,335

Extraversion

Kurzbezeichnung: D14, M14

Literatur Schupp, J., & Gerlitz, J.-Y. (2008). BFI-S: Big Five Inventory – SOEP. In A. Glockner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 12.00. Bonn: GESIS.

Anzahl Items: 3

Skalierung: 1=„trifft überhaupt nicht zu“; 7=„trifft voll zu“

Wie schätzen Sie Ihre/n Lehrer/in Ihres Deutschunterrichts/Mathematikunterrichts ein?

Variable		Skalenkennwerte			
Itemkürzel	Text	m	SD	rit	λ
<i>rec_D14a</i>	Unsere Lehrerin/ unser Lehrer ist zurückhaltend.	5,818	1,498	0,401	0,672
<i>rec_M14a</i>		5,633	1,597	0,401	0,660
D14b (Einsch_D14b)	Unsere Lehrerin/ unser Lehrer ist kommunikativ, gesprächig.	5,826	1,42	0,589	0,847
M14b (Einsch_M14b)		5,321	1,645	0,634	0,868
D14d	Unsere Lehrerin/ unser Lehrer geht aus sich heraus, ist gesellig.	5,222	1,564	0,618	0,865
M14d		4,869	1,708	0,684	0,894

Cronbach-Alpha (D) = 0,712

Cronbach-Alpha (M) = 0,741

Skalenmittelwert (D) = 5,621

Skalenmittelwert (M) = 5,271

ICC (D) = 0,270

ICC(M) = 0,404

Standardabweichung (D) = 1,194

Standardabweichung (M) = 1,339

Offenheit für neue Erfahrungen

Kurzbezeichnung: D14, M14

Literatur Schupp, J., & Gerlitz, J.-Y. (2008). BFI-S: Big Five Inventory – SOEP. In A. Glockner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 12.00. Bonn: GESIS.

Anzahl Items: 3

Skalierung: 1=„trifft überhaupt nicht zu“; 7=„trifft voll zu“

Wie schätzen Sie Ihre/n Lehrer/in Ihres Deutschunterrichts/Mathematikunterrichts ein?

Variable		Skalenkennwerte			
Itemkürzel	Text	m	SD	rit	λ
D14e	Unsere Lehrerin/ unser Lehrer hat eine lebhaft Phantasie, Vorstellung.	4,893	1,680	0,637	0,843
M14e		4,307	1,779	0,597	0,821
D14f	Unsere Lehrerin/ unser Lehrer schätzt künstlerische ästhetische Erfahrungen.	4,584	1,713	0,638	0,846
M14f		3,422	1,693	0,609	0,831
D14g	Unsere Lehrerin/ unser Lehrer ist originell, bringt neue Ideen rein.	4,785	1,758	0,610	0,825
M14g		4,170	1,839	0,617	0,834

Cronbach-Alpha (D) = 0,788

Cronbach-Alpha (M) = 0,773

Skalenmittelwert (D) = 4,750

Skalenmittelwert (M) = 3,967

ICC (D) = 0,246

ICC(M) = 0,292

Standardabweichung (D) = 1,439

Standardabweichung (M) = 1,470

Neurotizismus (gekürzt)

Kurzbezeichnung: D14, M14

Literatur Schupp, J., & Gerlitz, J.-Y. (2008). BFI-S: Big Five Inventory – SOEP. In A. Glockner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 12.00. Bonn: GESIS.

Anzahl Items: 2¹⁶³

Skalierung: 1=„trifft überhaupt nicht zu“; 7=„trifft voll zu“

Wie schätzen Sie Ihre/n Lehrer/in Ihres Deutschunterrichts/Mathematikunterricht ein?

Variable		Skalenkennwerte			
Itemkürzel	Text	m	SD	rit	λ
<i>rec_D14l</i>	<i>Unsere Lehrerin/unsere Lehrer wird leicht nervös.</i>	5,484	1,606	0,409	0,839
<i>rec_M14l</i>		5,343	1,722	0,488	0,863
D14k	Unsere Lehrerin/unsere Lehrer ist entspannt, kann mit Stress gut umgehen.	4,859	1,746	0,409	0,830
M14k		4,776	1,869	0,488	0,863

Cronbach-Alpha (D) = 0,579

Cronbach-Alpha (M) = 0,655

Skalenmittelwert (D) = 5,167

Skalenmittelwert (M) = 5,053

ICC(D) = 0,214

ICC(M) = 0,324

Standardabweichung (D) = 1,141

Standardabweichung (M) = 1,550

¹⁶³ Skala wurde gekürzt um ein Item: *“Unsere Lehrerin/unsere Lehrer macht sich oft Sorgen.“*

Verträglichkeit

Kurzbezeichnung: D14, M14

Literatur Schupp, J., & Gerlitz, J.-Y. (2008). BFI-S: Big Five Inventory – SOEP. In A. Glockner-Rist (Hrsg.), *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. ZIS Version 12.00. Bonn: GESIS.

Anzahl Items: 3

Skalierung: 1=„trifft überhaupt nicht zu“; 7=„trifft voll zu“

Wie schätzen Sie Ihre/n Lehrer/in Ihres Deutschunterrichts/Mathematikunterrichts ein?

Variable		Skalenkennwerte			
Itemkürzel	Text	m	SD	rit	λ
<i>rec_D14m</i>	<i>Unsere Lehrerin/ unser Lehrer ist manchmal etwas grob zu anderen.</i>	5,045	1,880	0,513	0,765
<i>rec_M14m</i>		4,975	1,928	0,551	0,782
D14n	Unsere Lehrerin/ unser Lehrer geht rücksichtsvoll und freundlich mit anderen um.	5,283	1,649	0,691	0,887
M14n		5,155	1,746	0,715	0,893
D14o	Unsere Lehrerin/ unser Lehrer kann verzeihen.	5,262	1,597	0,554	0,809
M14o		5,152	1,703	0,614	0,838

Cronbach-Alpha (D) = 0,752

Cronbach-Alpha (M) = 0,784

Skalenmittelwert (D) = 5,192

Skalenmittelwert (M) = 5,091

ICC (D) = 0,264

ICC(M) = 0,396

Standardabweichung (D) = 1,403

Standardabweichung (M) = 1,503

Prozessmerkmale des Unterrichts

Angemessenes Interaktionstempo

Kurzbezeichnung: D1, M1

Literatur: Clausen 2002, S.221: Interaktionstempo: BIJU (identisch).

Anzahl Items: 3

Skalierung: 1=„(fast) immer“; 2=„oft“; 3=„manchmal“; 4=„nie“

Wie denken Sie über das Unterrichtstempo und die Anforderungen in Ihrem Deutschunterricht/Mathematikunterricht?

Variable		Skalenkennwerte			
Itemkürzel	Text	M	SD	rit	λ
<i>rec_D1c</i>	<i>Unsere Lehrerin/ unser Lehrer verlangt oft blitzschnelle Antworten.</i>	3,119	0,823	0,548	0,824
<i>rec_M1c</i>		2,733	0,935	0,640	0,855
<i>rec_D1a</i>	<i>Unsere Lehrerin/ unser Lehrer lässt bei Fragen kaum Zeit zum Nachdenken.</i>	3,205	0,814	0,497	0,789
<i>rec_M1a</i>		2,864	0,947	0,605	0,833
<i>rec_D1g</i>	<i>Unsere Lehrerin/ unser Lehrer geht gleich zum Nächsten, wenn nicht sofort geantwortet wird.</i>	3,207	0,860	0,430	0,725
<i>rec_M1g</i>		2,982	0,927	0,534	0,779

Cronbach-Alpha (D) = 0,677

Cronbach-Alpha (M) = 0,761

Skalenmittelwert (D) = 3,176

Skalenmittelwert (M) = 2,859

ICC (D) = 0,144

ICC (M) = 0,248

Standardabweichung (D) = 0,649

Standardabweichung (M) = 0,770

Methodenvariation

Kurzbezeichnung: D3, M3

Literatur: Eigene Entwicklung.

Anzahl Items: 3

Skalierung: 1=„trifft nicht zu“/„nie“; 2=„trifft eher nicht zu“/„manchmal“; 3=„trifft eher zu“/„oft“; 4=„trifft zu“/„(fast) immer“

Wie denken Sie über die methodische Gestaltung des Unterrichts in Ihrem Deutschunterricht/ Mathematikunterricht?

Variable		Skalenkennwerte			
Itemkürzel	Text	m	SD	rit	λ
D3a	In diesem Unterricht findet häufig eine Variation der Methoden statt.	2,443	0,875	0,759	0,901
M3a		2,036	0,926	0,711	0,877
D3b	Dieser Unterricht ist methodisch abwechslungsreich.	2,505	0,895	0,768	0,907
M3b		2,087	0,932	0,768	0,909
<i>rec_D3c</i>	<i>In diesem Unterricht werden immer dieselben Methoden eingesetzt.</i>	2,637	0,935	0,636	0,823
<i>rec_M3c</i>		2,259	0,979	0,611	0,809

Cronbach-Alpha (D) = 0,849

Cronbach-Alpha (M) = 0,833

Skalenmittelwert (D) = 2,526

Skalenmittelwert (M) = 2,120

ICC (D) = 0,161

ICC (M) = 0,171

Standardabweichung (D) = 0,794

Standardabweichung (M) = 0,827

Disziplin

Kurzbezeichnung: D4, M4

Literatur: Waldis, Buff, Pauli und Reusser 2002; zit. n. Klieme, Pauli und Reusser 2005, S. 181, gekürzt Zeitverschwendung: Lange et al. 1983; Fend, 1977; Fend & Specht 1986 (id.), Clausen 2002, S.219, modifiziert, gekürzt Unterrichtsstörungen: BIJU, in Anlehnung an Eder (1985) (id.).

Anzahl Items: 4

Skalierung: 1=„(fast) immer“; 2=„oft“; 3=„manchmal“; 4=„nie“

Wie denken Sie über die Art und Weise, wie Ihre Klasse im Deutschunterricht/ Mathematikunterricht geführt wird?

Variable		Skalenkennwerte			
Itemkürzel	Text	m	SD	rit	λ
<i>rec_D4a</i> <i>rec_M4a</i>	<i>In diesem Unterricht dauert es meistens sehr lange, bis alle zur Arbeit bereit sind.</i>	2,688 2,683	0,877 0,919	0,618 0,675	0,800 0,831
<i>rec_D4d</i> <i>rec_M4d</i>	<i>In diesem Unterricht fehlt meistens bei irgendjemandem etwas, wenn wir anfangen wollen zu arbeiten.</i>	2,521 2,553	0,906 0,930	0,486 0,523	0,685 0,707
<i>rec_D4c</i> <i>rec_M4c</i>	<i>In diesem Unterricht wird oft sehr gestört.</i>	2,836 2,781	0,870 0,933	0,677 0,727	0,844 0,866
<i>rec_D4h</i> <i>rec_M4h</i>	<i>In diesem Unterricht wird fortwährend laut gequatscht.</i>	2,921 2,827	0,922 0,940	0,606 0,664	0,796 0,827

Cronbach-Alpha (D)=0,787

Cronbach-Alpha (M)= 0,822

Skalenmittelwert (D)= 2,742

Skalenmittelwert (M)= 2,710

ICC (D)=0,287

ICC(M)=0,331

Standardabweichung (D)= 0,698

Standardabweichung (M)= 0,753

Störungsprävention

Kurzbezeichnung: D4, M4

Literatur: Waldis, Buff, Pauli & Reusser 2002; zit. n. Klieme, Pauli & Reusser, 2005, S.182 CES (Moos/Trickett 1947) (id.), Waldis et al. 2002; zit. n. Klieme, Pauli und Reusser, 2005, S.192, gekürzt, mod. Störungspräventive Überwachung der Schülertätigkeiten: BIJU; Fend/Specht 1986 (1. Frage leicht modifiziert, 2. id.).

Anzahl Items: 4

Skalierung: 1=„trifft nicht zu“; 2=„trifft eher nicht zu“; 3=„trifft eher zu“; 4=„trifft zu“

Wie denken Sie über die Art und Weise, wie Ihre Klasse im Deutschunterricht/Mathematikunterricht geführt wird?

Variable		Skalenkennwerte			
Itemkürzel	Text	m	SD	rit	λ
D4e	In diesem Unterricht hat unsere Lehrerin/ unser Lehrer klargemacht, was passiert, wenn man (Verhaltens-)Regeln verletzt.	2,481	1,015	0,596	0,731
M4e		2,328	1,020	0,565	0,765
D4g	In diesem Unterricht wird klar, was man machen darf und was nicht.	2,731	0,923	0,707	0,764
M4g		2,672	0,948	0,692	0,752
D4b	Unsere Lehrerin/unser Lehrer merkt so- fort, wenn Schüler/innen beginnen, sich mit etwas anderem zu beschäftigen.	2,754	0,919	0,564	0,840
M4b		2,733	0,963	0,594	0,824
D4f	Unsere Lehrerin/unser Lehrer achtet sehr darauf, dass wir aufpassen.	2,879	0,900	0,706	0,854
M4f		2,870	0,938	0,677	0,845

Cronbach-Alpha (D)= 0,818

Cronbach-Alpha (M)= 0,812

Skalenmittelwert (D)= 2,705

Skalenmittelwert (M)= 2,643

ICC (D)=0,217

ICC(M)=0,254

Standardabweichung (D)= 0,749

Standardabweichung (M)= 0,770

Erklärkompetenz

Kurzbezeichnung: D5, M5

Literatur: Waldis, Buff, Pauli und Reusser 2002; zit. n. Klieme, Pauli und Reusser, 2005, S. 62, 2 Items gekürzt. Klarheit/Strukturiertheit des Unterrichts (2. Frage): BIJU; Fend/Specht, 1986; LASSO (id.), Waldis et al. 2002; zit. n. Klieme, Pauli und Reusser 2005, S. 63, gekürzt; modifiziert. Klarheit/Strukturiertheit des Unterrichts (1. Frage): BIJU; Fend & Specht, 1986; LASSO (id.).

Anzahl Items: 4

Skalierung: 1=„trifft nicht zu“; 2=„trifft eher nicht zu“; 3=„trifft eher zu“; 4=„trifft zu“

Wie denken Sie über die Klarheit und Struktur Ihres Deutschunterrichts/Mathematikunterrichts?

Variable		Skalenkennwerte			
Itemkürzel	Text	m	SD	rit	λ
D5a	Unsere Lehrerin/ unser Lehrer hebt immer wieder hervor, was wichtig ist.	2,976	0,832	0,533	0,723
M5a		2,987	0,945	0,581	0,739
D5d	Unsere Lehrerin/ unser Lehrer fasst häufig nochmals den Stoff zusammen, damit wir ihn uns gut merken können.	2,717	0,913	0,563	0,744
M5d		2,678	1,036	0,642	0,791
D5b	Unsere Lehrerin/ unser Lehrer kann gut erklären.	2,992	0,913	0,701	0,841
M5b		2,720	1,113	0,776	0,861
<i>rec_D5e</i>	<i>Unsere Lehrerin/ unser Lehrer erklärt so, dass keiner weiß, was los ist.</i>	3,252	0,908	0,593	0,776
<i>rec_M5e</i>		2,845	1,090	0,710	0,840

Cronbach-Alpha(D) = 0,788

Cronbach-Alpha (M) = 0,841

Skalenmittelwert (D) = 2,982

Skalenmittelwert (M) = 2,801

ICC (D) = 0,246

ICC (M) = 0,382

Standardabweichung (D) = 0,703

Standardabweichung (M) = 0,866

Motivierung

Kurzbezeichnung: D7, M7

Literatur Waldis, Buff, Pauli und Reusser 2002; zit. n. Klieme, Pauli und Reusser 2005, S.50. Motivierungsfähigkeit (Frage 1-3): BIJU (id.).

Anzahl Items: 3

Skalierung: 1=„trifft nicht zu“; 2=„trifft eher nicht zu“; 3=„trifft eher zu“; 4=„trifft zu“

Wie werden Sie im Deutschunterricht/Mathematikunterricht motiviert?

Variable		Skalenkennwerte			
Itemkürzel	Text	m	SD	rit	λ
D7a	Unsere Lehrerin/unsere Lehrerin gestaltet den Unterricht oft sehr spannend.	2,381	0,954	0,778	0,894
M7a		2,069	0,949	0,813	0,917
D7b	Unsere Lehrerin/unsere Lehrerin kann auch trockenen Stoff wirklich interessant machen.	2,383	0,998	0,803	0,907
M7b		2,100	0,953	0,828	0,928
D7d	Unsere Lehrerin/unsere Lehrerin kann Schüler/innen manchmal richtig begeistern.	2,386	0,963	0,711	0,848
M7d		2,154	0,992	0,761	0,889

Cronbach-Alpha (D)=0,877

Cronbach-Alpha (M)= 0,899

Skalenmittelwert (D)= 2,366

Skalenmittelwert (M)= 2,101

ICC (D)=0,231

ICC (M)=0,360

Standardabweichung (D)= 0,860

Standardabweichung (M)= 0,879

Alltagsbezug

Kurzbezeichnung: D7, M7

Literatur Waldis, Buff, Pauli und Reusser 2002; zit. n. Klieme, Pauli und Reusser 2005, S.57. Modifiziert, gekürzt.

Anzahl Items: 3

Skalierung: 1=„trifft nicht zu“; 2=„trifft eher nicht zu“; 3=„trifft eher zu“; 4=„trifft zu“

Wie werden Sie im Deutschunterricht/Mathematikunterricht motiviert?

Variable		Skalenkennwerte			
Itemkürzel	Text	m	SD	rit	λ
D7c	Unsere Lehrerin/unsere Lehrerin zeigt uns an Beispielen aus dem täglichen Leben, wozu man das Gelernte brauchen kann.	2,308	0,996	0,676	0,839
M7c		2,308	1,057	0,711	0,860
D7e	Wenn wir in diesem Unterricht etwas Neues erarbeiten, gehen wir meistens von unseren eigenen Erfahrungen und Alltagsbeispielen aus.	2,374	0,935	0,733	0,886
M7e		2,006	0,929	0,723	0,877
D7f	Um uns etwas zu erklären, nimmt unsere Lehrerin/unsere Lehrerin oft ein Beispiel aus dem täglichen Leben.	2,432	1,006	0,782	0,897
M7f		2,247	1,031	0,781	0,897

Cronbach-Alpha (D) = 0,856

Cronbach-Alpha (M) = 0,861

Skalenmittelwert (D) = 2,360

Skalenmittelwert (M) = 2,176

ICC (D) = 0,156

ICC (M) = 0,260

Standardabweichung (D) = 0,856

Standardabweichung (M) = 0,883

Lehrererhebung¹⁶⁴

Persönlichkeitsdimensionen

Gewissenhaftigkeit

Kurzbezeichnung: D_14, M_14

Literatur: Schupp, J., & Gerlitz, J.-Y. (2008). BFI-S: Big Five Inventory – SOEP. In A. Glockner-Rist (Hrsg.), Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version 12.00. Bonn: GESIS.

Anzahl Items: 3

Skalierung: 1=„trifft überhaupt nicht zu“; 7=„trifft voll zu“

Wie schätzen Sie sich persönlich ein?

Variable		Skalenkennwerte		
Itemkürzel	Text	m	SD	rit
rec_14c		6,022	1,209	0,410
recD_14c	<i>Ich bin ein Mensch, der eher faul ist.</i>	6,359	0,787	0,330
recM_14c		5,490	1,609	0,459
14h			5,993	0,920
D_14h	<i>Ich bin ein Mensch, der gründlich arbeitet.</i>	6,076	0,852	0,550
M_14h		5,898	1,085	0,671
14i			5,869	0,961
D_14i	<i>Ich bin ein Mensch, der Aufgaben wirksam und effizient erledigt.</i>	5,811	0,962	0,434
M_14i		5,837	0,986	0,562

Cronbach-Alpha = 0,659

Cronbach-Alpha (D) = 0,623

Cronbach-Alpha (M) = 0,711

Skalenmittelwert = 5,895

Skalenmittelwert (D) = 6,082

Skalenmittelwert (M) = 5,697

Standardabweichung = 0,881

Standardabweichung (D) = 0,657

Standardabweichung (M) = 1,040

¹⁶⁴ Itemkürzel ohne Fachangabe (D/M) stellen die fachübergreifenden Werte dar.

Extraversion

Kurzbezeichnung: D_14, M_14

Literatur: Schupp, J., & Gerlitz, J.-Y. (2008). BFI-S: Big Five Inventory – SOEP. In A. Glockner-Rist (Hrsg.), Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version 12.00. Bonn: GESIS.

Anzahl Items: 3

Skalierung: 1=„trifft überhaupt nicht zu“; 7=„trifft voll zu“

Wie schätzen Sie sich persönlich ein?

Variable		Skalenkennwerte		
Itemkürzel	Text	m	SD	rit
<i>rec_14a</i>		4,597	1,636	0,609
<i>recD_14a</i>	<i>Ich bin ein Mensch der zurückhaltend ist.</i>	5,019	1,380	0,252
<i>recM_14a</i>		4,122	1,889	0,590
14b		5,554	1,320	0,640
D_14b	Ich bin ein Mensch, der kommunikativ, gesprächig ist.	5,981	0,909	0,236
M_14b		5,061	1,519	0,748
14d		4,892	1,512	0,589
D_14d	Ich bin ein Mensch, der aus sich herausgeht, gesellig ist.	5,038	1,593	0,272
M_14d		4,571	1,486	0,621

Cronbach-Alpha = 0,773

Cronbach-Alpha (D) = 0,665

Cronbach-Alpha (M) = 0,798

Skalenmittelwert = 4,995

Skalenmittelwert (D) = 5,346

Skalenmittelwert (M) = 4,623

Standardabweichung = 1,267

Standardabweichung (D) = 1,025

Standardabweichung (M) = 1,398

Offenheit für neue Erfahrungen

Kurzbezeichnung: D_14, M_14

Literatur: Schupp, J., & Gerlitz, J.-Y. (2008). BFI-S: Big Five Inventory – SOEP. In A. Glockner-Rist (Hrsg.), Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version 12.00. Bonn: GESIS.

Anzahl Items: 3

Skalierung: 1=„trifft überhaupt nicht zu“; 7=„trifft voll zu“

Wie schätzen Sie sich persönlich ein?

Variable		Skalenkennwerte		
Itemkürzel	Text	m	SD	rit
14e	Ich bin ein Mensch, der eine lebhaftere Phantasie, Vorstellung hat.	5,06	1,468	0,525
D_14e		5,51	1,368	0,596
M_14e		4,80	1,262	0,329
14f	Ich bin ein Mensch, der künstlerische, ästhetische Erfahrungen schätzt.	5,06	1,786	0,473
D_14f		6,06	1,406	0,493
M_14f		4,24	1,880	0,287
14g	Ich bin ein Mensch, der originell ist, neue Ideen reinbringt.	4,99	1,252	0,603
D_14g		5,40	1,132	0,453
M_14g		4,62	1,260	0,576

Cronbach-Alpha = 0,702

Cronbach-Alpha (D) = 0,693

Cronbach-Alpha (M) = 0,561

Skalenmittelwert = 5,120

Skalenmittelwert (D) = 5,654

Skalenmittelwert (M) = 4,553

Standardabweichung = 1,191

Standardabweichung (D) = 1,029

Standardabweichung (M) = 1,091

Neurotizismus

Kurzbezeichnung: D_14, M_14

Literatur: Schupp, J., & Gerlitz, J.-Y. (2008). BFI-S: Big Five Inventory – SOEP. In A. Glockner-Rist (Hrsg.), Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version 12.00. Bonn: GESIS.

Anzahl Items: 3

Skalierung: 1=„trifft überhaupt nicht zu“; 7=„trifft voll zu“

Wie schätzen Sie sich persönlich ein?

Variable		Skalenkennwerte		
Itemkürzel	Text	m	SD	rit
<i>rec_14j</i>		3,921	1,720	0,584
<i>recD_14j</i>	<i>Ich bin ein Mensch, der sich oft Sorgen macht.</i>	3,462	1,720	0,597
<i>recM_14j</i>		4,120	1,792	0,662
<i>rec_14l</i>		4,971	1,424	0,602
<i>recD_14l</i>	<i>Ich bin ein Mensch, der leicht nervös wird.</i>	5,058	1,364	0,471
<i>recM_14l</i>		4,860	1,539	0,764
14k	Ich bin ein Mensch, der entspannt ist, mit Stress gut umgehen kann.	4,496	1,486	0,597
D_14k		4,558	1,349	0,620
M_14k		4,440	1,692	0,672

Cronbach-Alpha = 0,760

Cronbach-Alpha (D) = 0,730

Cronbach-Alpha (M) = 0,834

Skalenmittelwert = 4,401

Skalenmittelwert (D) = 4,333

Skalenmittelwert (M) = 4,473

Standardabweichung = 1,325

Standardabweichung (D) = 1,202

Standardabweichung (M) = 1,454

Verträglichkeit (gekürzt)

Kurzbezeichnung: D_14, M_14

Literatur: Schupp, J., & Gerlitz, J.-Y. (2008). BFI-S: Big Five Inventory – SOEP. In A. Glockner-Rist (Hrsg.), Zusammenstellung sozialwissenschaftlicher Items und Skalen. ZIS Version 12.00. Bonn: GESIS.

Anzahl Items: 2¹⁶⁵

Skalierung: 1=„trifft überhaupt nicht zu“; 7=„trifft voll zu“

Wie schätzen Sie sich persönlich ein?

Variable		Skalenkennwerte		
Itemkürzel	Text	m	SD	rit
14o		5,757	1,232	
D_14o	Ich bin ein Mensch, der verzeihen kann.	5,870	1,127	0,496
M_14o		5,640	1,336	
14n		5,767	0,899	0,556
D_14n	Ich bin ein Mensch, der rücksichtsvoll und freundlich mit anderen umgeht.	5,906	0,904	0,442
M_14n		5,620	0,879	

Cronbach-Alpha = 0,416

Cronbach-Alpha (D) = 0,048

Cronbach-Alpha (M) = 0,620

Skalenmittelwert = 5,762

Skalenmittelwert (D) = 5,887

Skalenmittelwert (M) = 5,6330

Standardabweichung = 0,857

Standardabweichung (D) = 0,732

Standardabweichung (M) = 0,963

¹⁶⁵ Skala wurde gekürzt um ein Item: „Ich bin ein Mensch, der manchmal etwas grob zu anderen ist.“

C: Faktorladungen der konfirmatorischen Faktorenanalysen der selbsteingeschätzten Lehrer- BigFive – entlang BFI-S/gekürzt (N = 103)

Ich bin ein Mensch, der	Gewissen- haftigkeit	Extra- version	Offenheit	Resilienz	Verträglichkeit
gründlich arbeitet.	,83**/,81**				
Aufgaben wirksam und effizient erledigt.	,66**/,71**				
<i>eher faul ist.</i>	,56**/,51**				
kommunikativ, gesprächig ist.		,83**/,87**			
aus sich herausgeht, gesellig ist.		,69**/,68**			
<i>zurückhaltend ist.</i>		,69**/,65**			
eine lebhafte Phantasie, Vorstellung hat.			,77**/,75**		
künstlerische ästhetische Erfahrungen schätzt.			,51**/,52**		
originell ist, neue Ideen reinbringt.			,71**/,73**		
entspannt ist, mit Stress gut umgehen kann.				,77**/,76**	
<i>leicht nervös wird.</i>				,69**/,68**	
<i>sich oft Sorgen macht.</i>				,77**/,76**	
rücksichtsvoll, freundlich mit anderen umgeht.					,44**/,55**
verzeihen kann.					,00/,50**
<i>manchmal etwas grob zu anderen ist.</i>					1,12**/ ---

D: Korrelationen zwischen Persönlichkeitswahrnehmungen nach Geschlecht

** bei Niveau 0,01 signifikant (fett gedruckt)/

* bei Niveau 0,05 signifikant (beides zweiseitig).

Korrelationen zwischen Schülerinnen und Lehrerinnen					
Lehrerinnen Schülerinnen	Gewissenhaftigkeit	Extraversion	Offenheit	Neurotizismus	Verträglichkeit
Gewissenhaftigkeit	,004	,089	-,050	-,014	,027
Extraversion	,111*	,171**	-,139**	,104*	-,004
Offenheit	,178**	,139**	,016	,066	,079
Neurotizismus	,083	,172**	-,081	,165**	-,031
Verträglichkeit	,084	,068	-,042	-,141**	,091

Anmerkung: Individualebene der Schülerinnen mit listenweisem Fallausschluss (N=433).

Korrelationen zwischen Schülern und Lehrern					
Lehrer Schüler	Gewissenhaftigkeit	Extraversion	Offenheit	Neurotizismus	Verträglichkeit
Gewissenhaftigkeit	,067	,163**	,179**	,229**	,072
Extraversion	-,072	,277**	,264**	,240**	-,014
Offenheit	,087*	,204**	,349**	,127**	,069
Neurotizismus	-,146**	,146**	,101*	,211**	-,018
Verträglichkeit	,035	,054	,188**	,172**	,102*

Anmerkung: Individualebene der Schüler mit listenweisem Fallausschluss (N=631).

Anhang

Korrelationen zwischen Schülern und Lehrerinnen					
Lehrerinnen Schüler	Gewissenhaftigkeit	Extraversion	Offenheit	Neurotizismus	Verträglichkeit
Gewissenhaftigkeit	-,011	-,182**	,071	-,108	-,036
Extraversion	,025	,018	,082	-,061	,027
Offenheit	,013	-,048	,198**	,000	,003
Neurotizismus	,133*	-,016	,094	,104	-,094
Verträglichkeit	,003	-,059	,090	-,134*	,084

Anmerkung: Individualebene der Schüler mit listenweisem Fallausschluss (N=328).

Korrelationen zwischen Schülerinnen und Lehrern					
Lehrer Schülerinnen	Gewissenhaftigkeit	Extraversion	Offenheit	Neurotizismus	Verträglichkeit
Gewissenhaftigkeit	,050*	,030	,077**	,056*	,024
Extraversion	-,003	,201**	,108**	,135**	-,009
Offenheit	,094**	,162**	,207**	,066**	,037
Neurotizismus	-,040	,059**	,015	,170**	-,058**
Verträglichkeit	,039	,024	,068**	,029	,058**

Anmerkung: Individualebene der Schülerinnen mit listenweisem Fallausschluss (N=2059).

Anhang

Korrelationen zwischen allen Schülerinnen/Schülern und Lehrerinnen					
Lehrerinnen Schüler und Schülerinnen	Gewissenhaftigkeit	Extraversion	Offenheit	Neurotizismus	Verträglichkeit
Gewissenhaftigkeit	,005	-,028	,001	-,048	-,002
Extraversion	,088*	,112**	-,051	,045	,006
Offenheit	,114**	,061	,091*	,048	,045
Neurotizismus	,115**	,095**	-,010	,144**	-,060
Verträglichkeit	,063	,021	,011	-,126**	,084*

Anmerkung: Individualebene der Schülerinnen und Schüler mit listenweisem Fallausschluss (N=766).

Korrelationen zwischen allen Schülerinnen/Schülern und Lehrern					
Lehrer Schüler und Schülerinnen	Gewissenhaftigkeit	Extraversion	Offenheit	Neurotizismus	Verträglichkeit
Gewissenhaftigkeit	,072*	,094**	,135**	,103**	,040
Extraversion	-,043	,264**	,197**	,185**	-,016
Offenheit	,087**	,232**	,277**	,075**	,035
Neurotizismus	-,113**	,109**	,057*	,155**	-,052
Verträglichkeit	,035	,027	,101**	,106**	,049

Anmerkung: Individualebene der Schülerinnen und Schüler mit listenweisem Fallausschluss (N=1277).

E1: Modell 1a – signifikante Regressionskoeffizienten der Lehrerpersönlichkeit auf die Unterrichtsqualität im saturierten Modell

Fach	Prädiktoren	Disziplin	Störungs- prävention	Erklär- kompetenz	Alltags- bezug	Interakti- onstempo	Motivie- rung	Methoden- variation
Schülererebene	Deutsch	Gewissenhaftigkeit	,11**	,23**	,09*		,12**	,10**
		Extraversion		,06*				
		Offenheit		,11**	,22**		,27**	,21**
		Resilienz	,09*					
		Verträglichkeit	,06*		,08**	,28**	,15**	,11**
	Erklärte Varianz (R ²)	,05**	,08**	,11**	,10**	,22**	,11**	
Schülererebene	Mathematik	Gewissenhaftigkeit	,09**	,20**			,07*	
		Extraversion						
		Offenheit		,13**	,23**		,27**	,16**
		Resilienz	,09**	,07*		,08**	,09**	
		Verträglichkeit	,09**	,19**	,09**	,30**	,14**	,08**
	Erklärte Varianz (R ²)	,05**	,09**	,09**	,14**	,20**	,06**	
Klassenebene	Deutsch	Gewissenhaftigkeit	,47**	,47**			,19*	,24**
		Extraversion						
		Offenheit						
		Resilienz	,53**	,26**	,30**	,46**	,28**	
		Verträglichkeit	-,34**	-,30**	,80**			
	Erklärte Varianz (R ²)	,61**	,85**	,50**	,87**	,80**	,46**	
Klassenebene	Mathematik	Gewissenhaftigkeit	,47**	,39**	,27**		,18**	
		Extraversion		,29**			,26**	-,32*
		Offenheit			,72**		,38**	,84**
		Resilienz	,58**	,17**		,86**	,21**	,34**
		Verträglichkeit	-,24**	-,75**				
	Erklärte Varianz (R ²)	,67**	,79**	,56**	,88**	,81**	,73**	

Anmerkung: Moderate und starke Effekte – gerundet $\geq 0,30$ – sind fett gekennzeichnet; auf Individual- ($N_{within} = 2247$) und Klassenebene ($N_{between} = 123$); Signifikanz_{within/between}: $p < 0,01/p < 0,05$.

E2: Modelle 2a/2b – signifikante Regressionskoeffizienten der Lehrerpersönlichkeit auf die Unterrichtsqualität – Korrelation zwischen Extraversion und Offenheit

Fach	Prädiktoren	Disziplin	Störungsprävention	Erklärkompetenz	Alltagsbezug	Interaktions-tempo	Motivierung	Methodenvariation	
Schülerenebene	Deutsch	Gewissenhaftigkeit	,11**/,13**	,23**/,26**	,09* /,10**		,13**/,15**	,11**/,11**	
		Extraversion		,06 ⁺					
		Offenheit		,11**/,12**	,15**/,16**	,22**/,23**		,28**/,29**	,21**/,20**
		Resilienz	,09**/,08*						
		Verträglichkeit	,06*/,07**		,18**/,20**	,08**/,09**	,28**/,31**	,16**/,17**	,11**/,12**
	Erklärte Varianz (R ²)	,05**/,05**	,05**/,05**	,18**/,18**	,08**/,08**	,09**/,09**	,16**/,16**	,08**/,08**	
Schülerenebene	Mathematik	Gewissenhaftigkeit	,09**/,13**	,18**/,17**	,20**/,21**		,07*/,08**		
		Extraversion							
		Offenheit		,09**/,13**	,14**/,17**	,23**/,25**		,28**/,30**	,16**/,23**
		Resilienz			,07* /n.s.		,08**/n.s.	,10**/n.s.	
		Verträglichkeit	,09**/,10**		,19**/,20**	,09**/,12**	,31**/,29**	,15**/,18**	,08**/,15**
	Erklärte Varianz (R ²)	,04**/,04**	,07**/,05**	,17**/,17**	,07**/,08**	,14**/,09**	,15**/,15**	,04**/,08**	
Klassenebene	Deutsch	Gewissenhaftigkeit	,48**/,41**	,78**/85**	,50**/,41**		,21 ⁺	,26**/,27**	
		Extraversion							
		Offenheit							
		Resilienz	,53**/,52**		,29**/,33**	,34**/,57**	,39**/,29**	,34**/,54**	,32**/,46**
		Verträglichkeit	,35**/,28**	,60**/,-42	,27**/,33**	,71**/,73**	,25*/,41**	,80**/,68**	,46**/,41**
	Erklärte Varianz (R ²)	,61**/,54**	,70**/,56**	,85**/,81**	,50**/,33**	,90**/,83**	,80**/,68**	,46**/,41**	
Klassenebene	Mathematik	Gewissenhaftigkeit	,46**/,48**	,50**/,82**	,46**/,61**	,31**/,36**		,22**/,25*	
		Extraversion						,32**/n.s.	-,34*/n.s.
		Offenheit		,63**/,30**	,34**/,33**		,77**/,43**	,44**/,61**	,93**/n.s.
		Resilienz	,58**/,58**		,20**/,24**			,26**/,29**	,38**/,32**
		Verträglichkeit	,28**/,42**	,35**/,-66**					
	Erklärte Varianz (R ²)	,65**/,62**	,69**/,67**	,69**/,72**	,43**/,31**	,87**/,73**	,72**/,63**	,63**/,25**	

Anmerkung: Moderate und starke Effekte – gerundet $\geq 0,30$ – sind fett gekennzeichnet; n.s. steht für nicht signifikant; $N_{within/between} = 2247/123$; Signifikanz_{within/between}: $p < 0,01 / < 0,05$;

F: Werte der Gesamtkonstrukte Persönlichkeit und Unterrichtsqualität auf Faktoren 2. Ordnung und ihre Regressionseffekte

Ladungen der schülerperzipierten BigFive-Dimensionen auf einem übergeordneten Faktor 2. Ordnung (als Gesamtkonstrukt Persönlichkeit):

BigFive-Dimension	Individual-/Klassenebene	
	Deutsch	Mathematik
Gewissenhaftigkeit	,69/,75	,64/,73
Extraversion	,63/,78	,62/,88
Offenheit für neue Erfahrungen	,61/,75	,54/,92
Resilienz	,55/,81	,57/,80
Verträglichkeit	,62/,70	,65/,72

Ladungen der schülerperzipierten Prozessmerkmale auf einem übergeordneten Faktor 2. Ordnung (als Gesamtkonstrukt Unterrichtsqualität):

Prozessmerkmal	Individual-/Klassenebene	
	Deutsch	Mathematik
Angemessene Methodenvariation	,51/,67	,42/,82
Motivierung	,76/,97	,74/,97
Angemessenes Interaktionstempo	,35/,66	,41/,74
Alltagsbezug	,56/,75	,53/,75
Erklärkompetenz	,74/,95	,74/,93
Störungsprävention	,52/,62	,45/,48
Disziplin	,39/,67	,39/,70

Effekte (Regressionskoeffizienten) der Gesamtkonstrukts Persönlichkeit auf das Gesamtkonstrukt Unterrichtsqualität:

Individual-/Klassenebene	
Deutsch	Mathematik
,68/,96	,68/,96
Erklärte Varianz	
,46/,93	,47/,92

Deutsch/Mathematik: RMSEA: ,06/,06; CFI: ,87/,86; TLI: ,84/,82

Lehrerpersönlichkeit und Unterrichtsqualität aus Schülerperspektive

Jessika Bertram

Die Lehrperson taucht in der Unterrichtsforschung vielfach als zentraler Parameter für gelungenen Unterricht auf, jedoch ohne empirische Evidenz. Während in diesem Kontext Schülerfeedbacks zur Unterrichtsqualität als Untersuchungsinstrument an Bedeutung gewinnen, erfolgt die Erfassung der Lehrerpersönlichkeit fast durchweg per Selbstbericht. Aufgrund von Erkenntnissen zu perspektivenspezifischer Validität von Feedbacks zur Unterrichtsqualität ist es jedoch plausibel, Lehrerpersönlichkeit und Unterrichtsqualität ebenfalls perspektivenspezifisch in Zusammenhang zu bringen. Die vorliegende Arbeit erfasst daher die BigFive der Lehrerpersönlichkeit sowohl aus Schüler- als auch aus Lehrkraftperspektive und darüber hinaus Schülerfeedbacks zu Prozessmerkmalen der Unterrichtsqualität. Korrelationsanalysen zwischen selbst- und fremd-wahrgenommener Lehrerpersönlichkeit zeigen, dass die verschiedenen Perspektiven differieren. Regressionsanalytische Befunde indizieren zudem, dass Einschätzungen von Lehrerpersönlichkeit und Unterrichtsqualität lediglich perspektivenspezifisch zusammenhängen.

25,00 €

ISBN 978-3-8405-0207-1

